

Exposé

zur Bachelorarbeit

Identifikation typischen Benutzerverhaltens in digitalen Studienformaten

im Studiengang

Medieninformatik Online

AUTOR

Werner Breitenstein
Matrikelnr.: 866059

BETREUER

Prof. Dr. Petra Sauer

ABGABE

25.02.2022

Situation

Studienabbrüche an Hochschulen verursachen Jahr für Jahr große finanzielle und gesellschaftliche Schäden und stehen daher schon seit geraumer Zeit auch im Fokus wissenschaftlicher Untersuchungen. Methoden des maschinellen Lernens (ML) sind hierbei von großer Bedeutung. Mit ihrer Hilfe lassen sich heute Modelle entwickeln, anhand derer die *Student Dropouts* mit einer recht hohen Wahrscheinlichkeit vorhergesagt werden können (Aulck, Nambi, Velagapudi, Blumenstock & West, 2019).

Standen in der Vergangenheit allein die traditionellen Studiengänge in Präsenzform im Interesse der Forschung, so gewinnen heute in zunehmendem Maße neue digitale Studienformate an Relevanz. Das von der Technischen Hochschule Lübeck und der Berliner Hochschule für Technik gemeinschaftlich initiierte Projekt *DiSEA – Digitale Studiengänge: Analyse von Erfolgs- und Abbruchfaktoren* nimmt diese neuen Entwicklungen in den Blick und adressiert hier insbesondere auch die Frage nach der Übertragbarkeit bisheriger Forschungsergebnisse aus den klassischen Studiengängen auf neue digitale Formate (Janneck, Merceron & Sauer, 2021).

Ein weiteres wichtiges Ziel von DiSEA besteht darin, die Vorhersagequalität der ML-Modelle anhand entsprechender Faktoren weiter zu verbessern. Dabei nehmen die Nutzungsdaten, die in Lernmanagementsystemen (LMS) von Studien- bzw. Kursbeginn an in hohem Umfang generiert werden eine immer bedeutsamere Rolle ein und motivieren damit genauere Untersuchungen (Janneck & Sauer, 2020).

Zielsetzung

Diese Arbeit, die sich nun im Rahmen von DiSEA mit der *Identifikation typischen Benutzerverhaltens in digitalen Studienformaten* beschäftigt, ist darauf gerichtet der zunehmenden Bedeutung des digitalen Lern- und Kommunikationsverhaltens an Hochschulen zu entsprechen (Baker, Lindrum, Lindrum & Perkowski, 2015). Auf Basis der Moodle-Daten zu Studiengängen der Virtuellen Fachhochschule (VFH) soll untersucht werden, ob und inwieweit sich typische Verhaltensweisen feststellen lassen und welche Daten bzw. Kennziffern hierzu sinnvoll zu berücksichtigen sind. Zusammenhänge mit Studienerfolgen oder sozio-demographischen Daten werden in dieser Arbeit nicht betrachtet.

Methodik

Den methodischen Schwerpunkt dieser Arbeit bildet die *Explorative Datenanalyse* (EDA), die auch einen festen Bestandteil im von Fayyad et al. (1996) entwickelten Knowledge Discovery in Databases-Prozess (KDD) darstellt (s. Abbildung 1). Mit Hilfe der EDA sollen die vorliegenden Daten der VFH eingehend statistisch untersucht und begutachtet werden, um z.B. Annahmen ableiten zu können, mögliche Querverbindungen zu entdecken oder auch Trends darzustellen.

Da die Daten selbst im wesentlichen die Vorgehensweise bestimmen, ist die im Vorfeld nach Auswahl der spezifischen Daten durchzuführende *Datenaufbereitung* von großem Gewicht. Dabei geht es nicht nur um die Korrektur möglicher Datenfehler in Form von Redundanz, Inkonsistenz oder falscher Formatierung, sondern auch um die Verbesserung der Datenqualität z.B. durch Vervollständigung lückenhafter Datensätze oder den Umgang mit sog. Ausreißern.

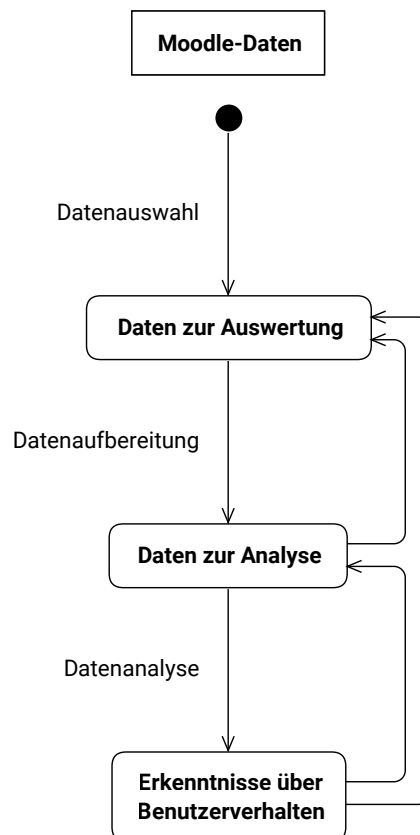


Abbildung 1: Phasen des (vereinfachten) KDD-Prozesses

Üblicherweise muss die Datenaufbereitung gemeinsam mit der sich anschließenden Datenanalyse in einem iterativen Prozess erfolgen (Beneker, 2020), um in jeder Phase der eigentlichen Analyse eine hochwertige und verlässliche Datenqualität zu garantieren.

Gegenstand der Untersuchungen zu dieser Arbeit ist ein durch die VFH zur Verfügung gestellter Datenbestand mit ca. 300.000 Datensätzen. In diesem enthalten sind die *Moodle-Nutzungsdaten aus vier Online-Studiengängen* aus dem WiSe2021¹:

1. Master-Studiengang Medieninformatik Online (MMIO)
2. Bachelor-Studiengang Wirtschaftsingenieurwesen Online (BWIO)
3. Bachelor-Studiengang Wirtschaftsinformatik Online (BWINF)
4. Bachelor-Studiengang Soziale Arbeit Online (BSAO)

Zur Durchführung der Datenanalyse kommen zunächst die folgenden beiden möglichen Vorgehensweisen in Betracht:

- *Zeitreihenanalyse*: Das Lern- und Kommunikationsverhalten wird hinsichtlich des Zeitpunkts und des Zeitraums in verschiedenen Auflösungen (zeitlichen Einheiten) und Ausdehnungen (zeitliche Dauer) untersucht.
- *Aktivitätsanalyse*: Gegenstand der Betrachtung sind Abfolgen von Aktivitäten, die typischerweise mehr oder minder häufig ausgeführt werden.

Im Zuge der Ausarbeitung können weitere Vorgehensweisen wie z.B. spezifische Mischformen hinzukommen, um die Signifikanz der Ergebnisse zu verbessern.

Die ermittelten Ergebnisse sind abschließend anschaulich zu visualisieren. Wie bei der Datenaufbereitung ist auch die Interpretation der gewonnenen Erkenntnisse und deren visuelle Darstellung als feste Bestandteile des Analyseprozesses wiederholt durchzuführen.

¹Das gesamte Semester musste nach der SARS-CoV-2-Infektionsschutzmaßnahmenverordnung des Berliner Senates unter erhöhten Sicherheitsbedingungen stattfinden. Die Regelungen für das Lehr- und Prüfungsgeschehen an der BHT wurden infolgedessen wie folgt angepasst:

- keine Lehrveranstaltungen und Prüfungen in Präsenz
- keine Zählung des Semesters als Fachsemester
- keine Zählung von Prüfungsfehlversuchen

Technische Grundlagen

Um in späteren Projekten nahtlos an die Erfahrungen und Erkenntnisse aus dieser Arbeit anknüpfen zu können, sollen in den verschiedenen Phasen der praktischen Untersuchungen nur quelloffene Standards zum Einsatz kommen. Im wesentlichen sind dies:

- *MySQL*: Der zu untersuchende Datenbestand der VFH, der initial in einem kommaseparierten CSV-Format bereitgestellt wird, wird zur weiteren Verwendung in eine MySQL-Datenbank migriert.
- *SQL*: Zu Beginn jeder neuen Analysephase sind die zu untersuchenden Daten mittels SQL-Anweisungen auszuwählen und aufzubereiten.
- *Python*: Aufgrund ihres großen Funktionsumfangs kommen bei der Datenanalyse verschiedene auf Python basierende Anwendungen und Bibliotheken zum Einsatz, z.B. Jupyter Notebook, NumPy, Pandas, Plotly oder Matplotlib.

Die Verwendung weiterer Tools wie z.B. auf Python basierende Frameworks zur Datenvisualisierung ist zum Zeitpunkt der Erstellung dieses Exposés nicht geplant. Über ihren möglichen Einsatz kann aus Zeitgründen erst im Zuge der eigentlichen Ausarbeitung entschieden werden.

Gliederung

1. Einleitung

- Motivation
- Zielsetzung
- Aufgabenstellung

2. Grundlagen

- Theorie
- Technik
- Datenbasis

3. Umsetzung

- Zeitreihenanalysen
 - Datenaufbereitung
 - Datenanalyse
 - Datenvisualisierung
- Aktivitätsanalysen
 - Datenaufbereitung
 - Datenanalyse
 - Datenvisualisierung

4. Ergebnisse

- Darstellung
- Diskussion

5. Fazit

6. Ausblick

Zeitplan

Zeitraum	Aufgaben
1. bis 4. Woche	Theoretisches Wissen aneignen Praktische Kompetenzen erarbeiten
1. bis 10. Woche	Literatur recherchieren und auswerten
3. bis 8. Woche	Analysen durchführen
9. bis 11. Woche	Bachelorarbeit erstellen
12. Woche	Bachelorarbeit korrigieren und abgeben

Während der Bearbeitungszeit werden u.a. zur Vorbereitung von Besprechungen kurze Statusberichte erstellt, die regelmäßig über den aktuellen Stand der Arbeiten und die geplanten weiteren Schritte informieren.

Literaturverzeichnis

- Aulck, L. S., Nambi, D., Velagapudi, N., Blumenstock, J. & West, J. D. (2019). Mining university registrar records to predict first-year undergraduate attrition. In *Edm.*
- Baker, R. S., Lindrum, D., Lindrum, M. J. & Perkowski, D. (2015). Analyzing Early At-Risk Factors in Higher Education e-Learning Courses. In *Proceedings of the Eighth International Conference on Educational Data Mining* (S. 150–155). International Educational Data Mining Society.
- Beneker, D. (2020). *Knowledge discovery in databases (kdd)*. Zugriff am 2022-02-24 auf <https://fh-bielefeld-mif-sw-engineerin.gitbooks.io/klausurthemen/content/ai/knowledge-discovery-in-databases.html>
- Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. (1996, Mar.). From data mining to knowledge discovery in databases. *AI Magazine*, 17 (3), 37. Zugriff auf <https://ojs.aaai.org/index.php/aimagazine/article/view/1230> doi: 10.1609/aimag.v17i3.1230
- Janneck, M., Merceron, A. & Sauer, P. (2021, March). Disea: Analysing success and dropout in online-degrees.. Zugriff auf <https://drive.google.com/drive/folders/1xXAKWkTlCeMb3PsmkiPEjNRq8G2pXelq>
- Janneck, M. & Sauer, P. (2020). *Analyse von Daten zum Lernverhalten*. Zugriff am 2022-02-24 auf <https://disea-projekt.de/index.php/ziele/>