

## Nutzung von Studienverlaufsdaten im Kontext eines Studienplanungsassistenten

### Erste Analysen und Limitationen

Tobias Hirmer<sup>1</sup>, Michaela Ochs<sup>1</sup>, Andreas Stöckl<sup>1</sup>, Adrian Völker<sup>1</sup>

**Abstract:** Die Analyse von Studierendenendaten bietet Potentiale für Hochschulen, Studienangebote zielgerichtet und studierendenfreundlich zu verbessern. Im Hinblick auf Studienverlaufsdaten muss jedoch deren Verfügbarkeit, Datenqualität und -quantität berücksichtigt werden. Der Beitrag zeigt erste Erfahrungen aus der Analyse von Studienverlaufsdaten von Studierenden ( $n = 181$ ) für die zukünftige Nutzung von maschinengestützten Verfahren. Es werden notwendige Transformationen und zu berücksichtigende Faktoren bei der Interpretation der Ergebnisse adressiert und die Implikationen für den Einsatz von KI-Systemen im Kontext eines Studienplanungsassistenten aufgezeigt.

**Keywords:** Studienplanung, Studienmonitoring, Studienverlaufsdaten, Kohortenanalyse

## 1 Einführung

Die Verbesserung von Studienangeboten und Studienbedingungen durch Universitäten stellt aus rechtlicher Perspektive [BA23] eine zentrale Aufgabe dar. Um entsprechende Verbesserungsbedarfe und Potentiale zu ermitteln, können mithilfe einer Analyse von bisher weitgehend ungenutzten Datenbeständen Handlungsbedarfe und -optionen für unterschiedliche Zielgruppen herausgearbeitet werden (siehe auch [Ta23, Wa23]). Dieser Beitrag soll die ersten Erfahrungen aus Projektvorhaben an der Universität Bamberg aufzeigen und Einblicke in derzeitige Exploration und die Nutzbarmachung von Studienverlaufsdaten für einen digitalen Studienplanungsassistenten (DSPA) aufzeigen. Hierzu werden der Kontext des DSPA kurz umrissen, die Erfahrungen insbesondere aus der statistischen Analyse erster Testdaten reflektiert und mögliche Ansatzpunkte für die Unterstützung durch KI-Systeme diskutiert. Aufgrund der Platzbeschränkungen in diesem Extended Abstract werden die Erkenntnisse hier ohne Bezug auf konkrete Beispiele beschrieben. Beispielhafte Modulkonstellationen werden im Workshop vorgestellt.

Das primäre Ziel des DSPA ist es, Studierende bei der Studienplanung zu unterstützen. Hierfür stellt das System auf Basis von Modulhandbüchern sowie einer Schnittstelle zum Lehrveranstaltungsverwaltungssystem grundlegende Funktionalitäten bereit. Diese umfassen ein Dashboard zur Abbildung des Studienverlaufs und -fortschritts, eine interaktive

---

<sup>1</sup> Universität Bamberg, Lehrstuhl für Medieninformatik, An der Weberei 5, 96047 Bamberg, {tobias.hirmer, michaela.ochs, andreas.stoeckl, adrian.voelker}@uni-bamberg.de

Darstellung des Modulangebots, eine Möglichkeit zur Stundenplanung für das aktuelle Semester sowie eine langfristige Planungskomponente für die Belegung von Modulen für kommende Semester. Gerade für letztere Planungsaufgabe bietet die Analyse von Studienverlaufsdaten<sup>2</sup> vielfältige Potenziale zur Entscheidungsunterstützung (siehe z. B. [Bi15]), indem die auf Basis von Kohorten ermittelten Erkenntnisse im Hinblick auf Belegungs- und Erfolgskorrelationen mit der studierendenspezifischen Belegungshistorie innerhalb des DSPA<sup>3</sup> für individuelle Empfehlungen kombiniert werden. Hierfür wurden erste Analysen<sup>4</sup> im Hinblick auf ein Modulempfehlungssystem zunächst auf der Datenbasis von Studienverläufen ( $n=181$ ) im Bereich der Informatik durchgeführt. Diese Daten wurden manuell aus dem Prüfungsverwaltungssystem extrahiert und pseudonymisiert, dienten bisher jedoch nur dem Zweck der Archivierung und Dokumentation.

## 2 Derzeitigen Explorationen

In einem ersten Schritt mussten die Testdaten in ihrer Struktur stark verändert werden, um sie für die Analyse vorzubereiten. Die Daten enthielten Noten Studierender, welche jedoch vor Kontextinformationen, wie dem Modul und dem Semester, in dem sie erzielt wurden, interpretiert werden müssen. Da diese Details in unserem Fall nur lose verknüpft waren, mussten die Informationen über Note, Modul und Semester für die Auswertung verbunden werden. Da jedes Modul in unterschiedlichen Semestern belegt werden konnte, mussten Datensätze angelegt werden, welche in der Lage waren, jede mögliche Kombination abzubilden und die klare Zuordnung von Modulnote, Modul und Semester zu erhalten. Für jeden Studierenden wurde ein einzelner Datensatz pro Semester angelegt, welcher alle belegbaren Module beinhaltet. Dies ermöglichte zwar die Auswertungen, erhöhte die Größe des Datensatzes jedoch substantiell. Derartige Lösungen sind jedoch davon abhängig, in welcher Form und Struktur die jeweiligen Studienverlaufsdaten vorliegen sowie von der Auswertungsmethode bzw. -software. Die benötigte Vorbereitung kann sich also in individuellen Fällen unterscheiden.

Eine weitere Herausforderung stellt die Individualität von Studienverläufen dar, insbesondere wenn der Studiengang viele Freiheiten in der Modulbelegung lässt. Eine Vielzahl individueller Verläufe wirkt sich ungünstig auf die Fallzahl spezifischer Fragestellungen aus und erschwert damit die statistische Auswertbarkeit. Eine beispielhafte Fragestellung lautet, ob sich das Bestehen eines Moduls günstig auf das Bestehen eines anderen Moduls auswirkt. Eine robuste Auswertung benötigt eine hinreichend große Anzahl von Studierenden, welche beide Module belegt haben, sowie auch eine Gruppe Studierender, welche nur eines der Module belegt hat, um einen robusten Vergleich zwischen diesen Gruppen zu ermöglichen. Anwendbar ist diese Fragestellung also primär auf Module, welche von

---

<sup>2</sup> Belegungszeitpunkt und -abfolge von Modulen und Modulnote

<sup>3</sup> Die individuelle Belegungshistorie wird derzeit aufgrund fehlender Schnittstellen zum Prüfungsverwaltungssystem explizit im DSPA erfasst.

<sup>4</sup> Diese beziehen sich auf die von der Universität bereitgestellten Studienverlaufsdaten. Die Interpretation von Analyseergebnissen sind zwangsläufig studiengang- bzw. universitätsspezifisch.

vielen Studierenden belegt werden. Bei Modulen mit kleinerer Belegung können mehrere dieser Module zu Gruppen inhaltsverwandter Module zusammengefasst werden, um für eine ausreichende Fallzahl zu sorgen. Jedoch muss hier auf die Strukturierung des Studiengangs geachtet werden. Müssen beispielsweise zwei bestimmte Module belegt werden, um das Studium erfolgreich abzuschließen, wird das Bestehen dieser Module mit hoher Wahrscheinlichkeit positiv korrelieren.

Da die zeitliche Abfolge der Modulbelegung variabel ist, ist die Frage der Kausalitätsrichtung von Relevanz. Variationen im Auswertungsmodell können sich also nicht nur auf die Module, sondern auch die zeitliche Abfolge beziehen. Zeigt sich eine positive Korrelation der Modulnoten von *Modul A* und *Modul B*, nur dann, wenn *Modul A* zuerst belegt wird, werden in *Modul A* vermutlich tatsächlich hilfreiche Kompetenzen für *Modul B* vermittelt. Eine Empfehlung, *Modul A* zuerst zu belegen, ist also sinnvoll. Korrelieren die Noten von *Modul A* und *B* auch, wenn *Modul B* zuerst belegt wird, werden womöglich in beiden Modulen ähnliche Kompetenzen vermittelt, deren Erwerb das Bestehen des anderen Fachs günstig beeinflusst. Viele Module können inhaltlich so konzipiert sein, dass sie ohne spezifisches Vorwissen verständlich sind. Eine andere mögliche Erklärung ist, dass eine günstige Prädisposition unter den Studierenden vorliegt, wie etwa Motivation oder Vorkenntnisse, welche erst zur Auswahl dieser Module geführt haben. Hier kann sich die Empfehlung lediglich darauf beziehen, dass beide Module inhaltlich verwandt scheinen, und eine Belegung beider Module aus einer erfolgsorientierten Perspektive sinnvoll ist.

Bei der Interpretation spielt auch das Semester eine Rolle. Je später ein Modul bestanden wird, umso größer ist die Anzahl der Module, die zuvor bestanden wurden, was die Korrelation zwischen diesen Modulen beeinflussen kann, sprich, der Erfolg später belegter Module wird eher mit dem Erfolg in früher belegten Modulen korrelieren. Ein tatsächlicher positiver Zusammenhang zwischen dem Bestehen von Modulen würde sich nur ergeben, wenn sich dieser auch bei der Belegung und dem Bestehen in früheren Semestern zeigt. Weiterhin muss die Anzahl der Versuche<sup>5</sup>, welche zum Bestehen eines Moduls benötigt werden, bei der Interpretation Berücksichtigung finden. Liegen mehrere Versuche für ein Modul vor, so wird dieses eher in späteren Semestern bestanden. Dies kann zu dem kontraintuitiven Ergebnis führen, dass das Bestehen eines früheren Moduls die Erfolgsaussichten eines späteren Moduls negativ beeinflusst. Dies ist jedoch unplausibel und in diesem Fall sollte eher von keinem Zusammenhang zwischen den Modulen ausgegangen werden. Eine Beschränkung auf Studierende, welche beide Module im Erstversuch bestanden haben, ist möglich, bringt jedoch das Problem der Selektivität mit sich.

Die bisherigen Analysen geben also Anlass zur Vorsicht bei der Interpretation von Ergebnissen. Notwendige Kontextdaten von Modulnoten, wie die Modulbezeichnung und das Semester, sowie die zeitliche Abfolge der Modulbelegung und die daraus resultierende Individualisierung von Studienverläufen sind Faktoren, welche bei der Konzeption und Interpretation von Auswertungen umfassend berücksichtigt werden müssen. Die zahlreichen Kombinationen machen eine manuelle Auswertung von Studienverlaufsdaten

---

<sup>5</sup> Versuch = Belegung eines Moduls. Die Regelung der Fehlversuche unterscheidet sich je nach Fakultät bzw. Universität. Hier sind unbegrenzt Versuche (im Rahmen der Höchststudiendauer) möglich.

schwierig. Letzte Limitationen verbleiben jedoch selbst bei umfassender Vorbereitung der Daten. Der Studienerfolg Studierender kann von zahlreichen weiteren Faktoren beeinflusst werden.<sup>6</sup> Es ist somit wichtig, neben der reinen Analyse von Studienverläufen weitere individuelle Faktoren bei der Generierung von Empfehlungen zu berücksichtigen. Zudem sollten Empfehlungen an Studierende wirklich nur als Empfehlungen gewertet und keinesfalls mit krisensicheren Erfolgsrezepten gleichgesetzt werden.

### 3 Diskussion und Ausblick

Die vorherigen Ausführungen sind besonders auch auf den geplanten Einsatz von Techniken aus dem Machine Learning (z. B. *kausale probabilistische Netze* [Fr13]) zu übertragen. Im Falle größerer Datenmengen ist eine automatisierte Lösung sowie die Aufteilung der Daten in Trainings-, Validierungs- und Evaluationsset im Kontext fundierter Empfehlungen sinnvoll, resultiert jedoch aufgrund der Individualität von Modulbelegungen bzw. Studienverläufen in der Regel in geringen Fallzahlen. Schwerpunkte der Implementierung solcher Ansätze im Kontext des DSPA sind somit der *Human in the loop* Ansatz (Unterstützung von KI-Systemen durch menschliche Komponente) sowie Aspekte der *Erklärbarkeit* (*Explainable AI*) zur Unterstützung von transparenten und informierten Entscheidungen seitens der Studierenden. Auch für weitere Zielgruppen bietet der DSPA und die Analyse von Studienverlaufsdaten Handlungsoptionen. So sind z. B. (Miss-)Erfolgsanalysen, insbesondere für das Studienmonitoring und die Verbesserung von Studienangeboten durch die Universität, relevant.

#### Literaturverzeichnis

- [BA23] BayHIG: Art. 82 Studienberatung, abgerufen 11. Mai 2023, <https://www.gesetze-bayern.de/Content/Document/BayHIG-82>.
- [Bi15] Bittner, P.: Modulempfehlungen über Vorgänger- und Nachfolgemodule. In (Cunningham, D. W., Hofstedt, P., Meer, K. & Schmitt, I., Hrsg.): INFORMATIK 2015, Bonn: Gesellschaft für Informatik e.V., S. 725-733, 2015.
- [Fr13] Friedman, N. et al.: Learning the Structure of Dynamic Probabilistic Networks. ArXiv preprint arXiv:1301.7374, 2013.
- [Ta23] Tauböck, S. et al.: PASSt – Predictive Analytics Services für Studienerfolgsmanagement. Zeitschrift für Hochschulentwicklung, 18 (Sonderheft Hochschullehre): S. 251–277, 2023.
- [Wa23] Wagner, M. et al.: A Combined Approach of Process Mining and Rule-Based AI for Study Planning and Monitoring in Higher Education. In (Montali, Marco; Senderovich, Arik; Weidlich, Matthias, Hrsg.): Process Mining Workshops. Lecture Notes in Business Information Processing, Springer Nature Switzerland, Cham, S. 513–525, 2023.

---

<sup>6</sup> z. B. persönliche Motivation, präexistente Kompetenzen oder Lernzeitrestriktionen durch Nebenbeschäftigungen