

Fragestellung und Ablauf des Themenblocks

Erarbeitet von
Dr. Katja Theune

Lernziele	1
Inhalt	2
Einstieg	2
Anwendungsfall für diesen Themenblock	2
Ablauf dieses Themenblocks	3
Quellen	3
Disclaimer	4

Lernziele

- Du kannst die Problemstellung dieses Themenblocks erläutern
- Du kannst Anwendungsbeispiele mit Bezug auf die Problemstellung nennen

Inhalt

Einstieg

Prognosen begegnen uns überall in unserem Alltag. Z. B. schauen wir täglich nach Wetterprognosen, bekommen Empfehlungen für Musik oder Serien und überlegen vielleicht, wie sich die Preise für Lebensmittel entwickeln werden.

Es geht bei Prognosen und Prognosemodellen einfach gesagt darum, mit verschiedenen Inputs, auch Features genannt, einen bestimmten Output bzw. eine Zielgröße vorherzusagen, und das möglichst genau. Z. B. werden uns mit verschiedenen Informationen über unser Sehverhalten möglichst passende Filmvorschläge gemacht. Für die Entwicklung von Prognosemodellen benötigen wir Verfahren des maschinellen Lernens. Wir befinden uns hier im überwachten Lernen, bzw. supervised learning, da die Modelle mit Daten trainiert werden, bei denen wir neben den Inputs auch die zugehörigen Outputs kennen. So können wir dann auch am Ende beurteilen, ob wir mit unserem Modell gute Prognosen machen.

Anwendungsfall für diesen Themenblock

Wir beginnen nun aber mit einem konkreten Anwendungsfall, der uns im weiteren Verlauf immer wieder zur Veranschaulichung der verschiedenen Themen begegnen wird. Und zwar werden an Hochschulen zunehmend sogenannte Frühwarnsysteme entwickelt, um z. B. abbruchgefährdete Studierende möglichst früh zu identifizieren. Ziel ist dabei, diese Studierenden dann mit individuellen Maßnahmen im Studium zu fördern, um einem möglichen Studienabbruch entgegenzuwirken. Diese Frühwarnsysteme basieren auf Prognosemodellen, die wiederum z. B. mit Verfahren des maschinellen Lernens entwickelt werden.

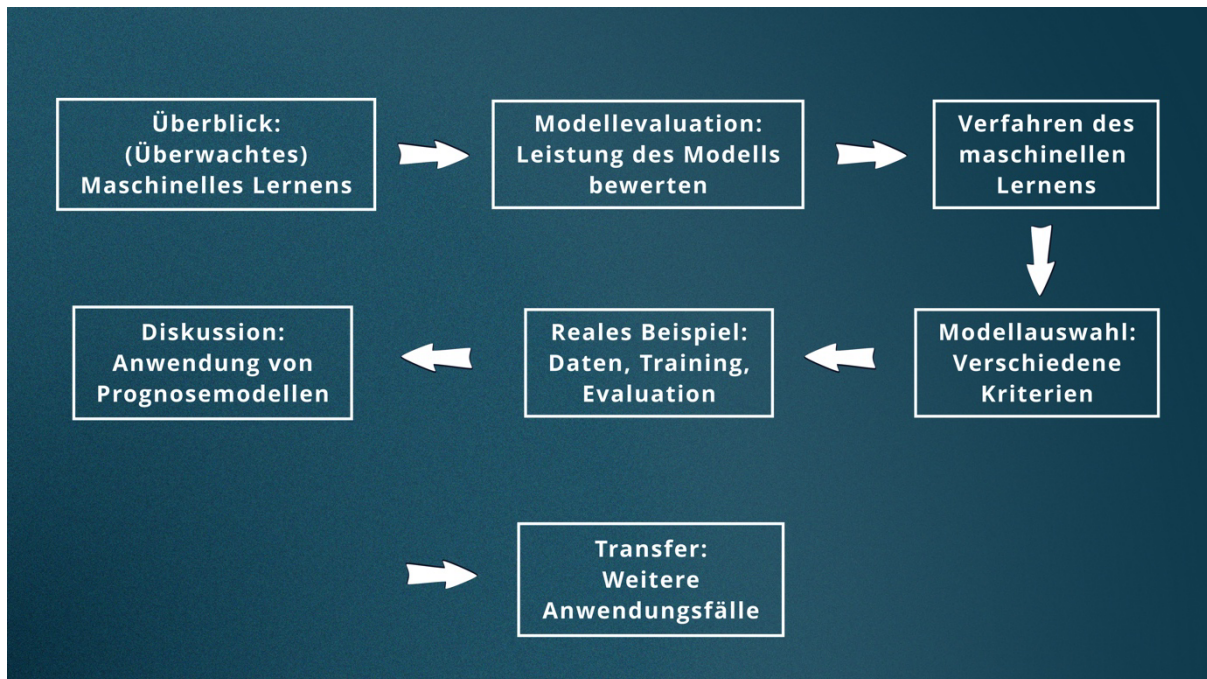
Quelle [1,2,3]

Wir können bestimmt alle gut nachvollziehen, dass für den Studienverlauf und -erfolg viele verschiedene Features bzw. Inputs eine Rolle spielen. Solche, die schon vor Studienbeginn als auch solche, die während des Studiums auftreten. Wir können uns hier z. B. den eigenen Bildungshintergrund, den Berufsstatus unserer Eltern, oder auch die Zufriedenheit mit dem Studium vorstellen.

Quelle [4]

Als Zielgröße bzw. Output wird häufig der Studienabbruch oder Studienerfolg verwendet. Hier befinden wir uns dann in der Klassifikation. Darauf wird auch unser Fokus liegen. Um uns aber auch Regressionsprobleme anzuschauen, greifen wir als Output dann auf die durchschnittliche Studienabschlussnote zurück.

Ablauf dieses Themenblocks



In diesem Themenblock beginnen wir mit einem kurzen Überblick über die verschiedenen Verfahren des maschinellen Lernens, hier insbesondere des überwachten Lernens, die wir für Prognosemodelle benötigen. Danach befassen wir uns damit, wie wir die Leistung unseres Prognosemodells bewerten bzw. evaluieren können. Daran anschließend gehen wir intensiver auf die einzelnen Verfahren des maschinellen Lernens ein und besprechen, wonach man eigentlich ein bestimmtes Verfahren für einen Anwendungsfall auswählt.

Damit es etwas praktischer wird, schauen wir uns das Ganze auch anhand eines realen Datensatzes beispielhaft an. Wir werden uns also die Daten genauer ansehen, auf diesen ein Prognosemodell trainieren und unsere Ergebnisse diskutieren. Wir besprechen dabei auch verschiedene weitere Aspekte, die beim Einsatz solcher Leistungsprognosen berücksichtigt werden sollten. Zuletzt schauen wir uns dann noch weitere mögliche Fragestellungen an, welche ganz ähnlich zu unserer Eingangsfragestellung bearbeitet werden könnten.

Quellen

- Quelle [1] Behr, A., Giese, M., Tegum K., H. D., & Theune, K. (2020). Early Prediction of University Dropouts – A Random Forest Approach. *Journal of Economics and Statistics*, 240 (6), 43-789. <https://doi.org/10.1515/jbnst-2019-0006>
- Quelle [2] Berens, J., Schneider, K., Gortz, S., Oster, S., & Burghoff, J. (2019). Early Detection of Students at Risk - Predicting Student Dropouts Using Administrative Student Data from German Universities and Machine Learning Methods. *Journal of Educational Data Mining*, 11(3), 1–41. <https://doi.org/10.5281/zenodo.3594771>

- Quelle [3] Theune, K. (2021). Determinanten und Modelle zur Prognose von Studienabbrüchen. In M. Neugebauer, H.D. Daniel, & A. Wolter (Hrsg.), Studienerfolg und Studienabbruch (S. 19-40). Springer VS, Wiesbaden. https://doi.org/10.1007/978-3-658-32892-4_2
- Quelle [4] Behr, A., Giese, M., Tegum K., H. D., & Theune, K. (2020). Dropping out of university: a literature review. Review of Education, 8 (2), 614-652. <https://doi.org/10.1002/rev3.3202>

Disclaimer

Transkript zu dem Video „Prognosemodelle (Klassifikation und Regression): Fragestellung und Ablauf des Themenblocks“, Dr. Katja Theune.

Dieses Transkript wurde im Rahmen des Projekts ai4all des Heine Center for Artificial Intelligence and Data Science (HeiCAD) an der Heinrich-Heine-Universität Düsseldorf unter der Creative Commons Lizenz [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/) veröffentlicht. Ausgenommen von der Lizenz sind die verwendeten Logos, alle in den Quellen ausgewiesenen Fremdmaterialien sowie alle als Quellen gekennzeichneten Elemente.