

Zusammenfassung

Daniel Siemmeister

Graz, 2022

Das Ziel dieser Arbeit ist, ausgehend von Daten von Studierenden, zu schätzen, wie viele Studierende in 3 Jahren prüfungsaktiv sein werden. Man will hierbei nur die Anzahl der prüfungsaktiven Studierenden vorhersagen, und nicht jeden einzelnen Studierenden individuell klassifizieren.

Um das Ziel der Arbeit zu erreichen, werden unterschiedliche Ansätze erprobt, welche auf verschiedenen Arten versuchen die Problemstellung zu lösen. Die Problemstellung wird in zwei kleinere Unterproblemstellungen unterteilt. Die erste der beiden beschäftigt sich mit der Vorhersage der prüfungsaktiven Studierenden, die bereits inskripiert sind. Die zweite Unterproblemstellung beschäftigt sich mit der Vorhersage von zukünftigen Studierenden, die in den darauffolgenden beiden Jahren inskripiert werden. In den meisten Ansätzen werden Machine Learning Modelle unterschiedlicher Architektur erprobt und hinsichtlich ihrer Vorhersagefähigkeit verglichen.

Die entscheidenden Einsichten der Arbeit sind:

- Verständnis der Problemstellung. Das heißt, es ist nur die Anzahl der prüfungsaktiven Studierenden gefragt, und man muss nicht jede Person exakt klassifizieren.
- Machine Learning Prädiktoren klassifizieren anhand von geschätzten Wahrscheinlichkeiten. Anstatt eines Schwellwertes zu klassifizieren, kann man diese Wahrscheinlichkeiten auch ohne Klassifizierung verwenden.
- Unterschiedliche Machine Learning Algorithmen liefern ähnliche Resultate. Die Komplexität der Algorithmen ist nicht entscheidend.
- Anhand von geschätzten Wahrscheinlichkeiten kann man die erwartete Anzahl an prüfungsaktiven Studierenden passabel vorhersagen.

Auch wenn die Vorhersagen gute Ergebnisse liefern, muss beachtet werden, dass man mehr Daten benötigt, um alle Ansätze hinreichend zu erproben.