

Abstract

Daniel Siemmeister

Graz, 2022

The aim of this paper is to estimate, based on student data, how many students will be active students in 3 years. The aim is only to predict the total number, and not every student individually.

To achieve the goal of the work, different approaches are tested, which try to solve the problem in different ways. The problem is divided into two smaller problems. The first of the two deals with the prediction of the number of students who are already enrolled. The second is concerned with predicting future students who will enrol in the following two years. In most of the approaches, machine learning models of different architectures are applied and compared in terms of their predictive ability.

The key insights of the work are:

- True understanding of the problem. That is, only the number of active students is required, and one does not need to classify each person exactly.
- Machine learning algorithms for classification give probabilities. Based on manually set thresholds, they try to classify. These probabilities can also be used without actual classification.
- Different machine learning algorithms produce similar results. The complexity of the algorithms is not decisive.
- Using estimated probabilities, one can reasonably predict the expected number of active students.

Although the predictions give good results, it is important to note that more data is needed to adequately test all approaches. In addition, there is a risk that there will be fundamental changes in student behaviour over the 3 years and that the predictions will not be realised. will not come true.