Predicting Student Performance using Machine Learning

Author: Lucas Jakin
Mentor: Branko Kavšek

The application of machine learning in educational settings has gathered significant attention, particularly in model predictions allowing businesses to make accurate guesses to the likely outcomes of a question based on historical data. My diploma thesis aims to analyze a dataset from Kaggle related to secondary school students in Portugal. By following the CRISP-DM methodology (Cross Industry Standard Process for Data Mining), I will systematically explore each phase, from data understanding and preparation to modeling and evaluation. In the modeling phase I will implement various ML algorithms, using different parameters every time, to predict students' academic success. The performance of these models will be then compared and tested to determine the most effective approach.

LITERATURE:

- https://www.kaggle.com/datasets/devansodariya/student-performance-data/data
- Cortez P., Silva A., 2008. "Using Data Mining to Predict Secondary School Student Performance". 4800-058. 8 pages
- Schröer C., Kruse F., Gomez J., 2021. "A Systematic Literature Review on Applying CRISP-DM Process Model". 526-534. 9 pages
- Chapman P., Kreber R., Clinton J., Khabaza T., Reinartz T., Wirth R., 1999. "The CRISP-DM Process Model". 99 pages.

- Oyeotun K., Tackie S., Olaniyi E., Khashman A., 2015. “Data Mining of Students’ Performance: Turkish Students as a Case Study”. 10 pages.