

Data Science Bootcamps

SDAIA Academy

❖ The Design

The project will be around education area where the goal of building the solution is to discover data about students' engagement and performance, The repository contains the datasets used as part of the OC2 lab's work on student Performance prediction and student engagement prediction in eLearning environments using machine learning methods [1].

Questions to be discovered:

- How students engagements in classes activities effect their overall engagement?
- How students grade on different activities effect their total grades?
- How students engagements in classes activities effect total grades?
- How to predict student who need support on early stages?

❖ The dataset

This repository contains the datasets used as part of the OC2 lab's work on Student Performance prediction and student engagement prediction in eLearning environments [1].

Student-Engagement-Prediction-eLearning-dataset

Feature	Engagement Metric Type	Description	Type	Value/s
Student Id		Student identifier	Nominal	std000, ..., std485
Number of Logins	Interaction	Number of times student accessed the course site on the LMS	Numeric	0, ..., 647
Number of Content Reads	Interaction	Number of times student accessed course material	Numeric	0, ..., 1007
Number of Forum Reads	Interaction	Number of times student read posts on the discussion forum	Numeric	0, ..., 58
Number of Forum Posts	Interaction	Number of times student posted on the discussion forum	Numeric	0, ..., 6
Number of Quiz Reviews	Interaction	Number of times student reviewed their quiz solution before final submission	Numeric	0, ..., 12
Assignment 1 lateness indicator	Effort	A binary indicator stating whether Assignment 1 submission is late or not	Numeric	0,1

Assignment 2 lateness indicator	Effort	A binary indicator stating whether Assignment 2 submission is late or not	Nu me ric	0,1
Assignment 3 lateness indicator	Effort	A binary indicator stating whether Assignment 3 submission is late or not	Nu me ric	0,1
Assignment 1 duration to submit (in hours)	Effort	The duration (in hours) between Assignment 1 posting and submission	Nu me ric	0, ..., 583
Assignment 2 duration to submit (in hours)	Effort	The duration (in hours) between Assignment 2 posting and submission	Nu me ric	0, ..., 297
Assignment 3 duration to submit (in hours)	Effort	The duration (in hours) between Assignment 3 posting and submission	Nu me ric	0, ..., 632
Average Assignment duration to submit (in hours)	Effort	The average duration (in hours) between Assignments' posting and submission	Nu me ric	0, ..., 496

Student-Performance-Prediction-eLearning-dataset

Feature	Description	Type	Value/s
Student Id	Student identifier	Nominal	std000, ..., std485
Quiz01	Quiz1 Mark	Numeric	0,...,100
Assign.01	Assign.01 Mark	Numeric	0,...,100
Midterm	Midterm Mark	Numeric	0,...,100
Assign.02	Assign.02 Mark	Numeric	0,...,100
Assign.03	Assign.03 Mark	Numeric	0,...,100
Final Exam	Final Exam Mark	Numeric	0,...,100
Final Grade	Total Final Mark	Numeric	0,...,100
Student Category	Final Grade	Nominal	G, F, W

❖ The algorithm

Given the requirement Multiclass classification algorithm seems to be suitable to be used on this project where to classify the students based on their engagement and permeance.

❖ The tools to be used in EDA, Modeling and Visualization

EDA will be used as follow:

- Pandas to importing dataset and to be used on manipulation operations such as merging, reshaping, selecting, as well as data cleaning, and data wrangling features.
- Seaborn and matplotlib to be used for data visualization and exploratory data analysis.

❖ References:

1. Abdallah Moubayed, MohammadNoor Injadat, Abdallah Shami, Ali Bou Nassif, Hanan Lutfiyya. (2020). Student Performance and Engagement Prediction in eLearning datasets. IEEE Dataport. <https://dx.doi.org/10.21227/4xkr-0f88>