Abstract

The project will be around education area where the goal of budling 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].

Design

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.

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?

data

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

Algorithm

Given the requirement binary classification algorithm and the model used is K- Means model, it seems to be suitable to this project where to classify the students based on their engagement and permeance.

Tools

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.

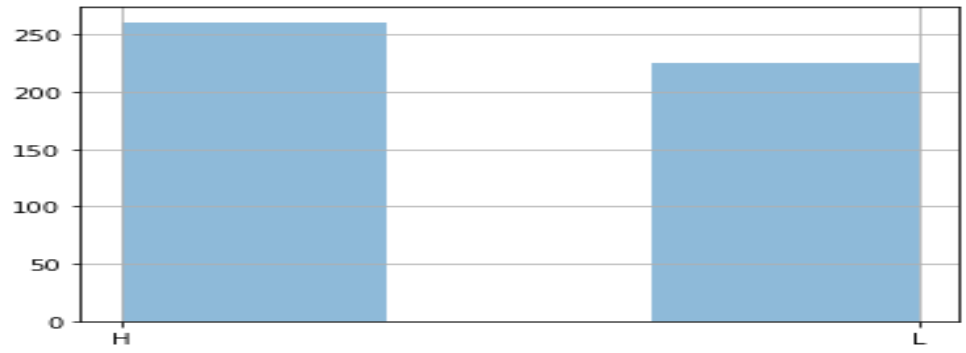
Communication

Classify students engagements activities

H - indicates high engagement

L – indicates low engagement

nt ID	# Logins	# Content Reads	# Forum Reads	# Forum Posts	# Quiz Reviews before submission	Assignment 1 lateness indicator	Assignment 2 lateness indicator	Assignment 3 lateness indicator	Assignment 1 duration to submit (in hours)	Assignment 2 duration to submit (in hours)	Assignment 3 duration to submit (in hours)	Average time to submit assignment (in hours)	Engagement_Level
0000	143	344	58	0	3	0	0	0	178.166667	92.716667	116.166667	129.016667	H
0001	70	342	0	0	4	0	0	0	294.033333	196.083333	217.750000	235.955556	L
0002	42	219	0	0	3	0	0	0	169.600000	235.733333	260.333333	221.888889	L
0003	92	271	2	0	6	0	0	0	341.150000	245.900000	271.216667	286.088889	L
0004	116	379	0	0	1	0	0	0	325.500000	236.283333	260.733333	274.172222	L
0005	62	358	0	0	0	0	0	0	182.800000	89.883333	107.666667	126.783333	H
0006	76	213	0	0	0	0	0	0	220.600000	200.483333	224.800000	215.294444	L
0007	78	237	0	0	3	0	0	0	171.216667	91.333333	111.800000	124.783333	H
0008	69	218	0	0	1	0	1	0	132.950000	296.000000	103.433333	177.461111	L
0009	89	360	9	0	3	0	0	0	304.933333	213.050000	232.666667	250.216667	L

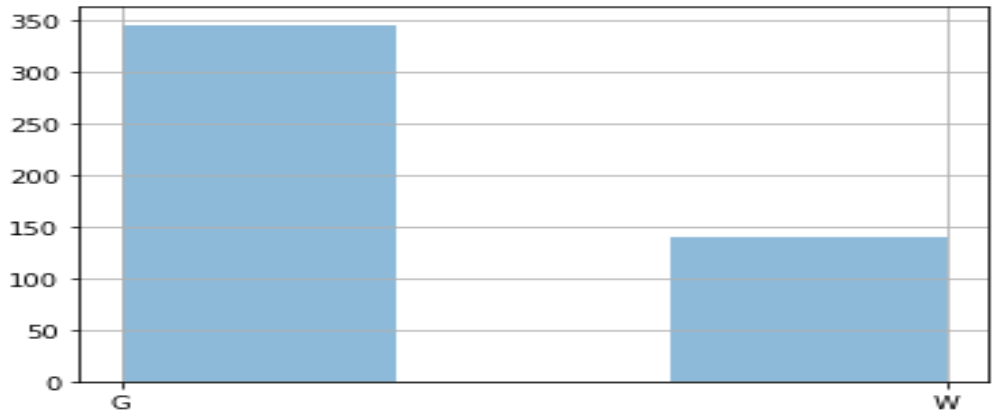


Classify students performance

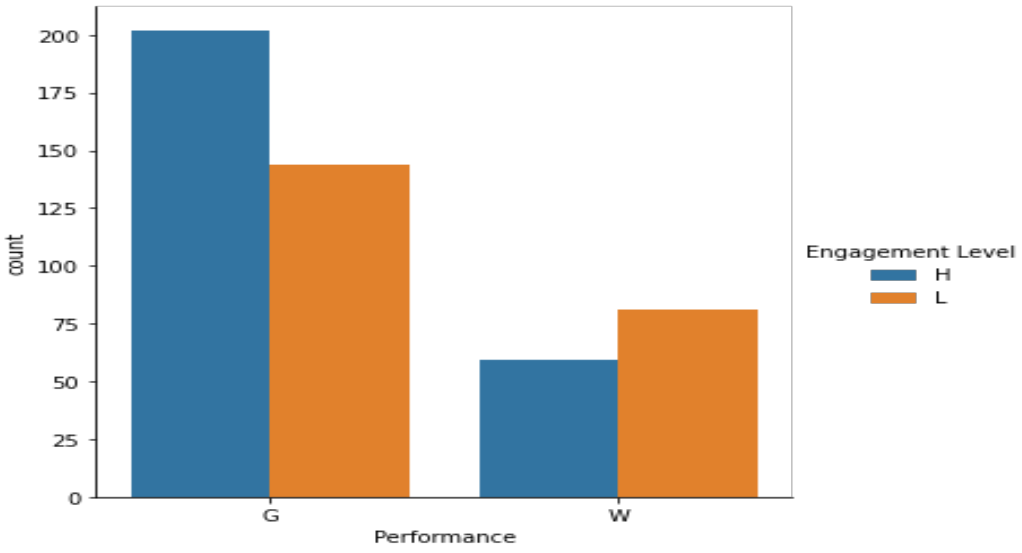
G - indicates good performance

W – indicates week performance

	Student ID	Quiz01 [10]	Assignment01 [8]	Midterm Exam [20]	Assignment02 [12]	Assignment03 [25]	Final Exam [35]	Course Grade	Total [100]	Performance
0	student000000	95	91	70	90	84	64	85	85	G
1	student000001	85	76	65	61	73	64	76	76	W
2	student000002	85	41	73	61	73	61	73	73	W
3	student000003	80	78	80	79	79	57	80	79	G
4	student000004	85	91	78	80	84	67	85	85	G
5	student000005	80	79	83	87	82	60	83	83	G
6	student000006	70	91	80	81	79	73	85	85	G
7	student000007	75	65	63	71	78	31	64	64	W
8	student000008	80	87	85	88	89	59	85	85	G
9	student000009	70	68	85	91	82	71	86	86	G



Observe the correlation between perfomance and engagement



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