Machine Learning Competition

- Team Darth Coders

Problem Statement

- In an alternate universe, due to the unbalanced workload among the faculty of Stanford University, the university is suffering from high faculty attrition and has become an object of mockery among the public. There has also been a petition to rename it to Standard University due to this mismanagement.
- Develop a model that tries to forecast the future total student enrolment for courses offered at the university based on the historic enrolment trend of the last 200 years.

Libraries Used

- Pandas
- Numpy
- Scikit-Learn
- Pytorch
- Fastai

Data Processing

- The team decided to approach the problem as a tabular dataset instead of a time series once.
- We made two extra columns Start and Stop from the "timestamp" column in the data set

The use of Fastai Library reduced our work to pre-process our data

Methods worked on

- We first started by giving the mean of the students of a particular course as our prediction. This established our **baseline model**
- We then used various regressors from the scikit learn library to model the data.
- We lastly implemented a Neural Network using the 'fastai' library to model the data. This used the categorical data as **embeddings instead of encoding it.**

Best Model

- The least RMSE values was achieved by the Random Forest Regressor with the number of trees set to 200

Learning Outcomes

- HackRush was a great experience to the team and we learnt a lot about ML.
- Since this was our first kaggle competition, we learnt a lot about new libraries.

Thank You

From team Darth Coders

Team Members:

1. Mihir Agarwal
2. Naman Dharmani
3. Shrey Joshi
4. Srujan Kumar Shetty