Description of the project

Project title: Student Dropout Prediction Through the Use of Machine Learning Method

Type of problem

Classification

Description of data set

- Source of data: https://www.kaggle.com/datasets/thedevastator/higher-education-predictors-of-student-retention?resource=download
  - This dataset provides a comprehensive view of students enrolled in various undergraduate degrees offered at a higher education institution. It includes demographic data, social-economic factors and academic performance information that can be used to analyze the possible predictors of student dropout and academic success. This dataset contains multiple disjoint databases consisting of relevant information available at the time of enrollment, such as application mode, marital status, course chosen and more. Additionally, this data can be used to estimate overall student performance at the end of each semester by assessing curricular units credited/enrolled/evaluated/approved as well as their respective grades.
- Response variable, Graduate, Dropout and Enrolled
- Number of observations,4424
- Number of predictors, 36.

https://figshare.com/articles/dataset/UnitelmaSapienza_1_0_zip/14554137

Description:
This is the dataset from the article "Hidden Space Deep Sequential Risk Prediction on Student Trajectories" by Bardh Prenkaj, Damiano Distante, Stefano Faralli and Paola Velardi

The document with detailed features information can be consulted at: http://valoriza.ipportalegre.pt/piaes/features-info-stats.html (accessed on 29 March 2023).

Statistical learning method(s) we want to use.

Prediction models proposed are.

- Artificial Neural Network (ANN),
- Decision
- Support Vector Machine
- Tree (DT) and
- Bayesian Networks **(BNs)**

To evaluate our models, we will use two different approaches. For dropout classification the data set is split in 70% train and 30% test, training the models using grid search and cross-validation on the training set and evaluating them on the test set.