



Contribution ID : 14

Type : **Invited talk**

Clustering educational data: a high school students' performance analysis

Tuesday, 31 January 2023 11:55 (25)

In this talk, we first briefly discuss the definitions of Educational Data Mining and Learning Analytics, providing an overview of the most used statistical methods in both fields, as well as a synopsis of their differences and similarities. The possible aims of the different methods, aimed at uncovering hidden patterns in educational data, are stressed pointing to the possible services provided to the whole school community. In the following, we focus on clustering methods for educational data: we start from traditional methods (hierarchical, partitive, and density-based), we proceed with dimension reduction methods, such as factorial k-means and reduced k-means, and we present some methods which incorporate the longitudinal dimension. Then, we present a pilot analysis carried out on a dataset reporting the performance of a class of high school students in three periods (which were treated as three separate datasets), using hierarchical clustering, partitive clustering (k-means), factorial k-means and reduced k-means techniques. The goal of the analysis is to show how the composition of groups and the number of groups vary in each period and which are the factors that influence the creation of groups. The partitions obtained with these algorithms were compared in terms of reliability using the average silhouette width index. Reduced k-means and k-means generated similar results and we can say that these results were the most acceptable considering the average silhouette width. Hierarchical clustering generated the same results as the former algorithms only in the first two periods of time. The results generated by factorial k-means differ from the other methods and as suggested by the values of the average silhouette width, it is not the best algorithm for clustering on the dataset available to us. The underlying meaning of clusters over time and the reasons behind statistical results are discussed and analyzed in detail, with the aim to highlight possible student group structures present in a high school class.

Research Strand

Data Science for Learning Processes and Education

Primary author(s) : FARNÈ, Matteo (University of Bologna)

Co-author(s) : Mrs. TARABORRELLI, Gioia

Presenter(s) : FARNÈ, Matteo (University of Bologna)

Session Classification : Assessing student competence: current methodological issues in case studies (Invited Symposium)

Track Classification : Assessing student competence: current methodological issues in case studies (Invited Symposium)