

Efficient use of big identifiable data for monitoring in education

Supervisor: P. Fazendeiro Co-supervisor: P. Prata

1 Abstract

Despite the increasing availability of data brought by technological advances, there have been very limited educational policies and programs that have capitalized on the evidence obtained from the big data analyses. Several authors include the administrative data as Big Identifiable Data (BID) source. Our aim is to contribute to increase the reliability of performance indicators (PIs) for monitoring in education by the efficient use of big data. Based on BID analytics, this research project aims to provide each Agrupamento de Escolas (AE, schools cluster) the software EduBigger to produce education PIs, for monitoring and evaluation purposes, contributing to the improvement of decision-making within the scope of school self-evaluation and public school advance, reinforcing the autonomy of the AEs.

We will use the AE gathered data as a way to improve the quality of data imputation, allowing the indicators to be based on all students data. It is expected that a better imputation enables better decision-making and produces an effective reinforcement of the predictive power of the synthesized tools. The large volume of data gathered by every AE, within the scope of the school self-evaluation, is unstructured in the sense that, although it presents some structure or may even be highly structured, its structure emerges in ways that are unanticipated. Thus the final objective of this work is to develop data-driven imputation methods using machine learning. These are based on the construction of a predictive model to estimate absent values from the information available in the data set.

2 Objectives and Tasks

- T1** Review the related literature
- T2** Gather the datasets of BID
- T3** Collect the requirements and define system's architecture
- T4** Computation of performance indicators
- T5** Construction of a predictive model for data imputation
- T6** Implement, test and deploy the application
- T7** Write up a MSc thesis and a scientific paper

3 Timetable

T1 1 month

T2 1 month

T3 1 month

T4 1 month

T5 1 month

T6 2 months

T7 2 months

4 Expected Results

- 1 journal paper
- 1 MSc thesis

5 References

M. E. Ferrao, G. Barros, A. M. Bof, and A. S. Oliveira, Estudo longitudinal sobre eficacia educacional no Brasil: Comparacao entre resultados contextualizados e valor acrescentado, Dados - Rev. Ciencias Sociais, vol. 61, no. 4, pp. 265300, 2018.

P. Fazendeiro and P. Prata, Performance assessment of the canonical genetic algorithm: A study on parallel processing via GPU architecture, in Evolutionary Computation. Apple Academic Press, 2017, pp. 325344.

P. Fazendeiro and J. V. de Oliveira, Observer-biased fuzzy clustering. IEEE Transactions on fuzzy systems, IEEE Trans. Fuzzy Syst., vol. 23, no. 1, pp. 8597, 2015.