

COVER LETTER OF THE KDMILE EXTENDED ARTICLE FOR PUBLISHING IN THE JOURNAL JIDM

Campina Grande - PB, July 15th 2016.

TO THE EDITORIAL COMMITTEE,

Dear Editors,

In response to the JIDM reviewers' requests and suggestions, the extended version of our article published at KDMiLe'15 entitled "*Exploiting Academic Records for Predicting Student Dropout: A Case Study in Brazilian Higher Education*" includes the new material listed below. Each item represents the main suggestions by each reviewer.

- Reviewer B suggestions:

1) *Give a better explanation about what we mean by "not considering" in the abstract;*

We rewrote the abstract sentence mentioned by Reviewer B to "We contrast both models and show that we have better results by not distinguishing the students by course.";

2) *Give a better justification for choosing Computer Science and Pharmacy as examples in section 5.2.1;*

We added a sentence explaining why we chose Computer Science and Pharmacy. In summary, they are chosen for illustration purposes since it is unfeasible to show all the courses. Nevertheless, the conclusions drawn for these two courses hold for the other courses as well;

3) *Figure 1 – a cumulative distribution function graph - may be interpreted as an increase of dropout instead of a decrease;*

4) *Table 1 - table showing the cost and number of dropouts per semester of course - could be substituted by a bar graph.*

In order to clarify these points we replaced Figure 1 for an image containing two bar graphs representing the number of drop outs per semester and the cost per semester respectively.

5) *English. Please consider send it to a proofreading service. There are many direct Portuguese translations found in the text.*

We have reviewed the whole text again and improved the writing.

- Reviewer C suggestions:

1) *The authors did not mention the tool used for all the experiments. Please explain it;*

We have Added a footnote mentioning the packages and tools used in our experiments.

2) *The year is lacking in some references;*

Completed the references with the missing years;

3) *Figure 3 - Science Computation -> Computer Science.*

Done

4) *The image definition of Figure 3 is not ok, and also this figure is not well explained. What means "gain ratio - oversample", and so on?;*

5) *Other points concerning the lack of details for how we used oversampling techniques in our experiments.*

We recreated Figures 3 and 4 (now they are Figure 4 and Figure 5), relabeled the models configurations described on them and added a new Section (sec. 5.1) to better explain the experiments (showing the factors considered and possible configurations) and how to identify each configurations described in the Figures.

Best regards,

Allan Sales

Leandro Balby

Adalberto Cajueiro