

Higher educational institutes generate massive amounts of student data.

REVIEW

Author page – No comments. Just try to check out if there are not middle names, which should be abbreviated as Initials .

Work structure – This work is well structured, including whole PeerJ standard sections.

Abstract -

- a. Reason of the work – Understanding various facets of student. Student segmentation framework with prediction model to provide a useful mechanism for devising pedagogical policies to increase the quality of education.
- b. Methodology – Educational data mining approach classification and mitigating academic failure and encouraging higher performance.
- c. Results – Experimental results indicate the effectiveness of the proposed framework and the applicability of classifying students into multiple performance levels using a small subset of courses being taught in the initial 2 years of the 4-year degree program.

Introduction -

The field (Educational Data Mining) is well described and the the need of this review is explained.

The relevance to readers in the field, and associated, is explained.

Previous reviews are cited.

Lines 84, 86 - You should provide your sections' numbers.

Related works -

Ligne 92 – Nghe et al., 2007 . **Check this.**

Ligne 150 - It is apparent that there is no 'best' classification algorithm, as different classifiers have outperformed each other in the discussed papers. **I think you should precise here that it is based on your literature exploration.**

Feature selection -

Ligne 275 – It is not always enough to have vast amounts of student data across multiple attributes to build effective classification models; **I don't agree with this statement, as an effective model depends on the quality of training (more you have data, better is your training). Can you review it?**

Discussion -

The current research validates that it is possible to successfully predict student performance at the end of the degree using student data at some earlier point during the courses of the degree.

Conclusion and future work -

3 questions were examined:

- The first question is focused on the generation of a classification model for early identification of student end-of-degree performance using the most basic and readily available learning data collected by higher educational institutes.
- The second question is focused on deriving courses that strongly influence the final prediction of student performance.
- The third question involved the generation of a segmentation framework.

The resultant segmentation matrix identifies students in various performance segments. The early identification of these students provides the opportunity to robustly devise a pragmatic policy to specifically target each performance level.

The future direction is given.

An important future direction can be to explore student performance in these courses. This will provide the educational institute an added opportunity to improve educational outcomes. Also, using the approach outlined in this paper, predictive models can be built for the early identification of student performance across the other degree programs offered by the university.

LACKS OF THE WORK -

- **The lack on the methodology, as for other previous works, is the absence of a method to early anticipate the selection of the best classifier. This should be done at the phase of the model evaluation and should be considered in the future work.**
- **The size of the data samples used (291, if compared for example with 20.492 used in a previous work). As you can see, if compared with other references, the size is too small, although it does not impact the overall outcomes of this work.**
- **The formats of the references of the literature resources, which do not respect the standards from PeerJ. Then the following references should be reformulated:**

Lines: 33, 38, 42, 46, 47, 49, 50, 53, 55, 56, 61, 62, 63, 64, 65, 67, 68, 89, 108, 113, 121, 141, 158, 159, 161, 162, 166, 168, 170, 171, 173, 177, 180, 182, 184, 187, 189, 193, 198, 202, 205, 211, 214, 217, 218, 219, 222, 226, 238, 241, 243, 245, 266, 271, 277, 279, 287, 288, 326, 397, 398, 419.