

2021

The Elusive Metrics - Are We Telling the Full Story in Educational Data Mining?

Keith Quille

Keith Nolan

Stephen Colgan

Follow this and additional works at: <https://arrow.tudublin.ie/ittscicon>

This Conference Paper is brought to you for free and open access by the School of Science and Computing at ARROW@TU Dublin. It has been accepted for inclusion in Conference Papers by an authorized administrator of ARROW@TU Dublin. For more information, please contact arrow.admin@tudublin.ie, aisling.coyne@tudublin.ie, gerard.connolly@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-NonCommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

The Elusive Metrics - Are We Telling the Full Story in Educational Data Mining?

Keith Quille
TU Dublin
Ireland
keith.quille@tudublin.ie

Keith Nolan
TU Dublin
Ireland
keith.nolan@tudublin.ie

Stephen Colgan
TU Dublin
Ireland
x00151280@mytudublin.ie

ABSTRACT

The use of Education Data Mining (EDM) has seen a significant increase in recent years. A recent report identified notable concerns with the literature relating to the lack of metrics presented in EDM research (in particular, predicting student performance). This poster presents details on these concerns that may inhibit future re-validation studies or worse, models that initially report strong findings which may not generalise. This poster also declares a call to action for future studies to present such metrics, and finally describes ongoing work in this space (a systematic literature review).

KEYWORDS

EDM, Educational Data Mining, Metrics, Re-validation

ACM Reference Format:

Keith Quille, Keith Nolan, and Stephen Colgan. 2021. The Elusive Metrics - Are We Telling the Full Story in Educational Data Mining?. In *26th ACM Conference on Innovation and Technology in Computer Science Education V. 2 (ITiCSE 2021), June 26–July 1, 2021, Virtual Event, Germany*. ACM, New York, NY, USA, 1 page. <https://doi.org/10.1145/3456565.3460071>

1 MOTIVATION

Computer Science Educational (CSEd) as a research area has exploded in recent years with new tools and pedagogical approaches being developed, all in the name of improving Computer Science attrition rates. A relevant review by Quille [2, 3] as part of his PhD thesis in 2018 examined the literature specifically to predicting student performance in CS1. The thesis took a systematic approach and identified several areas of significant concern when reporting prediction models in EDM. 1,884 articles were reviewed, resulting in 93 articles shortlisted with 49 articles included in the final review. One of the main findings (while there were multiple concerns) was that prediction models are often presented with a high accuracy, but that alone does not always reveal the entire story of the model. If sensitivity and specificity are not presented and identified, two concerns can be raised. First, a biased model could have been constructed. For example, if 90% of the students in a study were strong (high performing), the model could predict every student as strong and report an accuracy of 90%. This model presents as very

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

ITiCSE 2021, June 26–July 1, 2021, Virtual Event, Germany

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8397-4/21/06.

<https://doi.org/10.1145/3456565.3460071>

successful, but its ability to identify weak students (the model's main goal) would in-fact be 0% as it predicted every student as strong, thus the 10% of students that are weak, were incorrectly identified. Second, the accuracy does not present outcomes for a particular class, therefore making appraisals of models that aim to predict students who are struggling, masked when the study only presents accuracy. Only 6% of studies examined reported sensitivity or specificity (in some cases indirectly, but it could be calculated). This makes for appraisal and re-validation of such models difficult.

2 CURRENT WORK

Following on from the work of Quille, a systematic literature review (SLR) is now being conducted, where an initial investigation of 999 papers resulted in the inclusion of 86 papers in the review. Thematic analysis resulted in four major categories: Student Factors, Student Experience/Behaviour, Tools and Pedagogy. This SLR continues the work of Quille, with early indicators suggesting a significant increase in EDM studies conducted in recent years. This work is yet to be unpacked, however, early findings show a raise in the number of EDM papers published since Quille's work concluded.

3 A CALL TO THE COMMUNITY

A 2015 ITiCSE Working Group, Ihantola et al. [1] published five Grand Challenges relating to Educational Data Mining. Their fifth Grand Challenge was for future authors to "generalize the results to other contexts, if possible, and help practitioners apply them in their respective fields". Given the plethora of research in this field there is an absolute need for authors to be cognisant when presenting metrics. As this discipline is quite young, there is perhaps an element of unknown comprehension and a lack of awareness of suitable metrics to present in this space. Our call to the community is when presenting EDM models, to present as many relevant metrics as possible, some of which may include: 1) accuracy, 2) sensitivity and specificity, 3) class break down, 4) attributes used, 5) attribute selection techniques, 6) data normalisation techniques and 7) model hyper-parameters and topologies.

REFERENCES

- [1] P. Ihantola, A. Vihavainen, A. Ahadi, M. Butler, J. Börstler, S. H. Edwards, E. Isohanni, A. Korhonen, A. Petersen, K. Rivers, M. A. Rubio, J. Sheard, B. Skupas, J. Spacco, C. Szabo, and D. Toll. Educational data mining and learning analytics in programming: Literature review and case studies. In *Proceedings of the 2015 ITiCSE on Working Group Reports, ITiCSE-WGR '15*, pages 41–63, New York, NY, USA, 2015. ACM.
- [2] K. Quille. *Predicting and Improving Performance on Introductory Programming Courses (CS1)*. PhD thesis, National University of Ireland Maynooth, Maynooth University, 2019.
- [3] K. Quille and S. Bergin. CS1: how will they do? How can we help? A decade of research and practice research and practice. *Computer Science Education*, 29:254–282, 2019.