

Literature review on data modelling patterns and techniques for educational data assessment



Vilnius
universitetas

Snieguolė Bagočienė
Institute of Data Science and Digital Technologies
Vilnius University



Introduction

Educational data modelling plays a crucial role in developing theoretical frameworks, creating predictive models, and establishing guidelines for educational data collection and analysis. This poster presentation delves into the literature surrounding data modelling patterns and techniques applied to the assessment of educational data. As education increasingly relies on data-driven decision-making, this review explores scientific contributions that enhance our understanding of this vital field.

Our study comprehensively reviews and summarizes key concepts, theories, and methodologies found in existing literature. We showcase various data modelling models discussed in academic research and their applications in predicting student achievement, identifying learning trends, and adapting educational practices.

Additionally, the scholarly debate on the significance of data quality assurance and pre-processing in maintaining accuracy and reliability in educational data models is explored. Following this literature review, we outline best practices and emerging trends in educational data modelling.

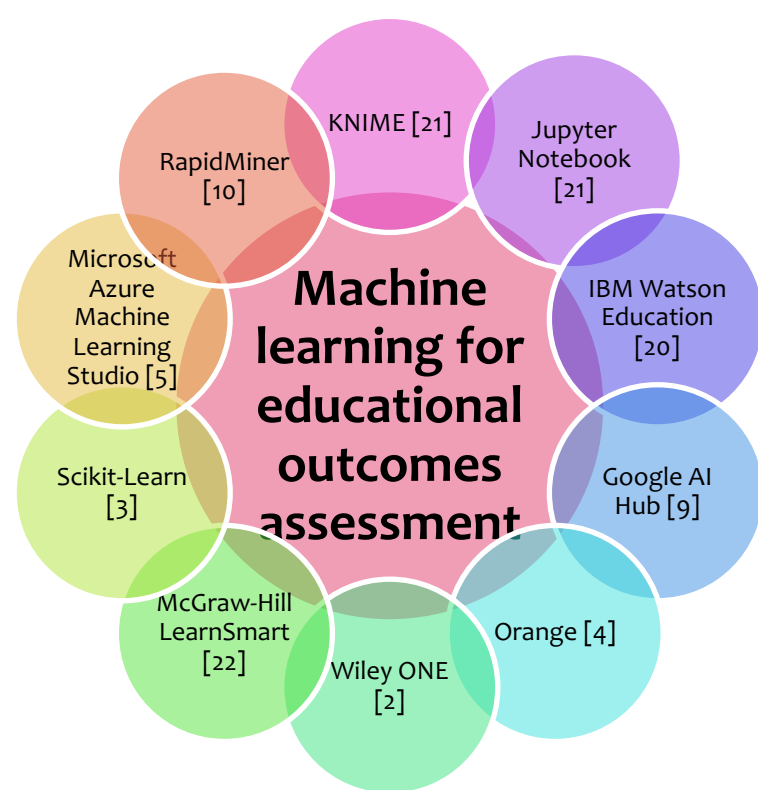
Methods

Literature review on:

The specific methods or techniques for educational outcomes assessment.

Results

Data Modeling Patterns and Techniques [7]



Assessment system is being developed to facilitate the work of teachers, consisting of the following items [12, 16]:

- Needs Assessment – identify teacher challenges and institution requirements.
- Define objectives – clearly outline assessment system goals and success criteria.
- Data collection and analysis – gather and analyze relevant student performance data.
- Model design – design a machine learning-based assessment model.
- Feature selection – prioritize relevant assessment features for analysis.
- Algorithm development – develop and train tailored machine learning algorithms.
- User interface design – create an intuitive interface for teacher interaction.
- Testing and validation – rigorously test and validate the system's accuracy.
- Implementation and integration – integrate the system into existing infrastructure with teacher training.
- Continuous improvement – monitor, gather feedback, and update the system as needed.
- Ethical considerations – address ethical concerns for fairness and student privacy.

What are some possible data modelling models or approaches that are commonly used in education data management?

The review talks about the need to change how we teach because there's a lot more data in education now. It mentions big data platforms and programs like MapReduce that help handle this big amount of data quickly. It also talks about a new way of teaching called CaMeLOT that follows Bloom's ideas. CaMeLOT helps teach how to model data in a clearer way for software engineering students. It's a structured way of learning that can make handling educational data better [6, 12]. Fink's model reflects a broader view of learning, emphasising not only the acquisition of knowledge and skills, but also the development of a holistic understanding that encompasses personal and interpersonal aspects. As education continues to evolve, frameworks such as the Fink model help to continuously explore effective teaching and learning strategies that meet the diverse needs of education [14]. DW solutions have also been proposed in the field of secondary education to identify strengths and weaknesses in educational environments. Although big data in education research has focused on modelling and DW design, there is a gap in the literature for a holistic methodology that integrates the entire educational data mining (EDM) process [15].

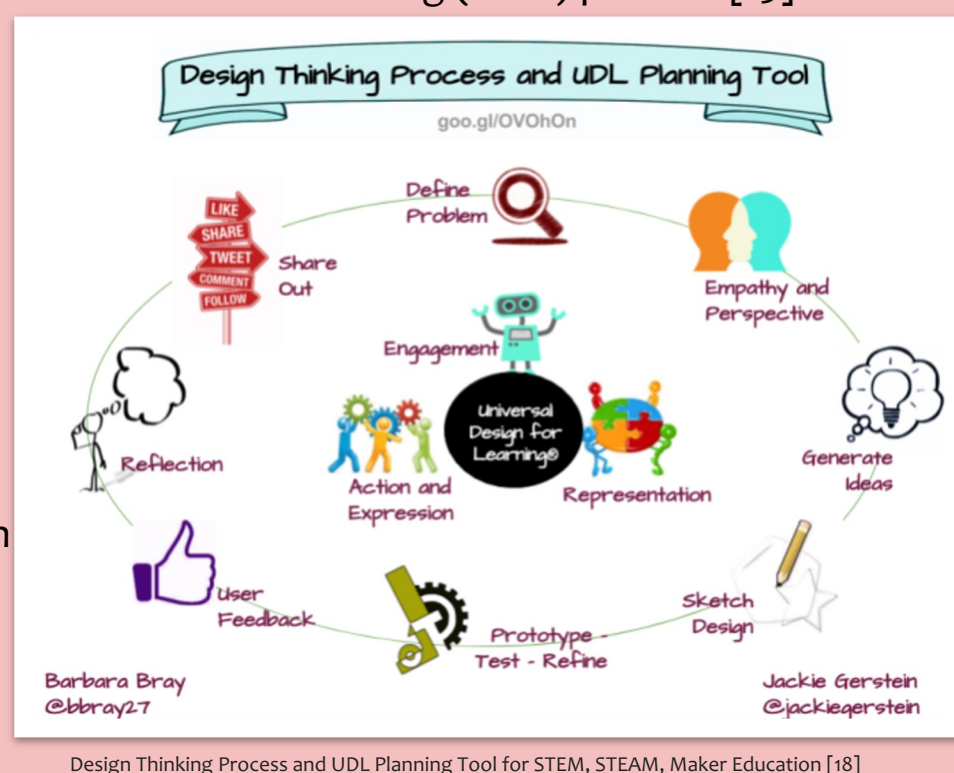
Steps in Data Modeling

- The process for data modeling [8]:
- defining entities and key properties,
- identifying relationships,
- mapping properties to entities,
- reducing redundancy, and
- completing and validating the data model.

This systematic approach ensures a thorough understanding of data requirements and relationships.

Data Modeling Tools

- ER/Studio, DbSchema, HeidiSQL, Toad Data Modeler, and ERBuilder [8].



What are some prevalent dimensional modelling patterns applicable to scenarios involving large-scale datasets [17, 19]

Star Schema

Snowflake Schema

Galaxy Schema

Data Lake

Normalization

This technique involves organizing data in a way that reduces redundancy and dependency. In education data modeling, normalization can help ensure that information is stored efficiently and consistently. For example, student information might be stored in one table, and course information in another, linked by unique identifiers.

Hierarchical data model

This model organizes data into a tree-like structure. In education, this could be applied to represent the hierarchy of educational institutions, with levels such as district, school, department, and class.

Data warehousing

Creating a centralized repository (data warehouse) for educational data allows for efficient reporting and analysis. It involves extracting, transforming, and loading (ETL) data from various sources into a structured format for analysis and reporting purposes.

Star schema and snowflake schema

These are schemas commonly used in data warehousing. The star schema has a central fact table connected to dimension tables, while the snowflake schema extends this concept by normalizing dimension tables. These schemas can be applied to educational data for analytics and reporting purposes.

Temporal data modeling

Education data often involves tracking changes over time, such as student enrollment, grades, and teacher assignments. Temporal data modeling techniques, like adding effective date ranges to records, can be useful to capture historical information.

Object-Oriented data modeling

This approach represents entities as objects, with attributes and methods. While more common in software development, it can be applied in educational data modeling, especially in systems where a more object-oriented approach is beneficial.

Graph data model

Graph databases can be useful for modeling relationships between entities in education, such as connections between students, teachers, and courses. This is particularly relevant for representing complex relationships and dependencies.

Metadata modeling

Including metadata in the data model is crucial for understanding the context and meaning of the data. In education data, metadata could include information about the source of the data, data quality, and any transformations applied.

Security and Privacy considerations

Incorporating data security and privacy features into the data model is essential, especially in educational systems handling sensitive student and personnel information.

Conclusions

Educational data mining and modelling, though distinct, are crucial for improving education [11]. Integrating these approaches yields valuable insights and supports personalized instruction and informed decision-making. Challenges include adapting to evolving educational environments and addressing ethical considerations [1].

In the field of education, insights are derived from large amounts and variety of real-time data, known as big data [6]. In order to speed up the analysis of educational data, the use of different big data platforms and parallel programming models could be explored for various data analysis techniques. When designing a data model for education, it's crucial to consider the **specific requirements, goals, and characteristics** of the educational institution or system in question. The chosen model should support efficient data storage, retrieval, and analysis while accommodating future changes and expansions.

Collaboration between education specialists and data scientists is seen as crucial for the success of EDM processes [15].

The landscape of educational data modelling is continually evolving, driven by technological advancements and a deeper understanding of educational processes [4, 5]. Notable emerging trends include the **integration of machine learning** algorithms for more accurate predictive modelling, the application of **natural language processing** techniques to analyse unstructured educational data, and the development of **adaptive modelling frameworks** that respond to individualized learning needs. These trends reflect an ongoing commitment to enhancing the effectiveness of educational data modelling in addressing the complexities of the educational landscape.

References

1. Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. Office of Educational Technology, US Department of Education. Dulhare.
2. Ahmad, K., & Bin Ahmad, K. A. (2020). Machine Learning and Big Data - Concepts, Algorithms, Tools and Applications (1st ed.). John Wiley & Sons. <https://doi.org/10.1002/9781119654834>
3. Géron. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. O'Reilly.
4. Godec, Pančur, M., Ilenič, N., Copar, A., Stržar, M., Erjavec, A., Pretnar, A., Demšar, J., Starič, A., Toplak, M., Žagar, L., Hartman, J., Wang, H., Bellazzi, R., Petrović, U., Garagna, S., Zuccotti, M., Park, D., Shauly, G., & Zupan, B. (2019). Democratizing image analytics by visual programming through integration of deep models and small-scale machine learning. Nature Communications, 10(1), 4551–4557. <https://doi.org/10.1038/s41467-019-12397-x>
5. Elati, L., & Elati, L. (2019). Azure machine learning studio. Machine Learning with Microsoft Technologies: Selecting the Right Architecture and Tools for Your Project, 201-223.
6. Wasson. (2015). Discovering Big Data Modelling for Educational World. Procedia, Social and Behavioral Sciences, 176, 642-649. <https://doi.org/10.1016/j.sbspro.2015.01.522>
7. What is Data Modeling? Types, Steps and Tools. (2023) <https://www.xenonstack.com/insights/data-modelling>
8. Dabney, & Eid, F. (2023). Beyond Bloom's: Fink's taxonomy as a catalyst for meaningful learning in nursing education. Teaching and Learning in Nursing. <https://doi.org/10.1016/j.teln.2023.09.007>
9. Luan, & Tsai, C.-C. (2021). A Review of Using Machine Learning Approaches for Precision Education. Educational Technology & Society, 24(1), 250-266. https://doi.org/10.30101/ETS.202101_241.0019
10. Kurasova. (2014). Programinės sistemos duomenų tyrybos mokymui. Lietuvos matematikos rinkinys. Ser. B, 55, 60-65. https://doi.org/10.15388/LMR_b.2014.12
11. Dol, & Jawandhiya, P. M. (2023). Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining — A survey. Engineering Applications of Artificial Intelligence, 122, 106071. <https://doi.org/10.1016/j.engappai.2023.106071>
12. Maki, P. L. (2011). Assessing for Learning: Building a Sustainable Commitment Across the Institution (2nd ed.). Routledge. <https://doi.org/10.4324/9781003443056>
13. Bogdanova, & Sneck, M. (2019). CaMeLOT: An educational framework for conceptual data modelling. Information and Software Technology, 110, 92-107. <https://doi.org/10.1016/j.infsof.2019.02.006>
14. Granzetti, J., Gisondi, M.A., Hopson, L.R., Regan, L. Aiming Beyond Competent: The Application of the Taxonomy of Significant Learning to Medical Education. Teach Learn Med. 2019 Aug-Sep;31(4):466-478. [Abstract]
15. 16- 15 Rozita Tsoni, Georgia Garani & Vassilios S. Vergyios (2023) Data pipelines for educational data mining in distance education, Interactive Learning Environments, DOB 10.1080/10494820.2022.2160466
16. Tai, J., Ajawi, R., Boud, D., Dawson, P., & Panadero, E. (2018). Developing evaluative judgement: enabling students to make decisions about the quality of work. Higher education, 75, 467-481.
17. Interneto nuododa. (2023.11) <https://medium.com/@sekindin/introduction-to-dimensional-data-modeling-c4be095eed1>
18. Post by Jackie Gerstein, Ed.D. @jackiegerstein and Barbara Bray @bbray27. Crossed posted at <http://barbarabray.net/2017/06/08/design-thinking-process-and-udl-planning-tool/>
19. Nonleco, A. (2020). Learn Data Warehousing in 24 Hours. Guru99.
20. Hoyt, R. E., Snider, D. H., Thompson, C. J., & Mantravadi, S. (2016). IBM Watson analytics: automating visualization, descriptive, and predictive statistics. JMIR public health and surveillance, 2(2), e5810.
21. Kathuria, A., Gupta, A., & Singla, R. K. (2021). A review of tools and techniques for preprocessing of textual data. Computational Methods and Data Engineering: Proceedings of IEMDE 2020, Volume 1, 407-422.
22. Sun, Q., Abdourazakou, Y., & Norman, T. J. (2017). LearnSmart, adaptive teaching, and student learning effectiveness: An empirical investigation. Journal of Education for Business, 92(1), 36-43.