The Impact of Data Feature Selection on AI-Based Student Performance Prediction: An Analysis of Educational Use Cases

Dr. Mahmoud Yousef AlFaress[1] 0009-0001-8949-024X, Prof. Dr. Midhunchakkaravarthy Janarthanan[2] 0000-0002-0107-885X, Prof. Dr. Chandra Kumar Dixit [3]

[1,2] Lincoln University College – Malaysia; [3] Institute of Engineering and Technology, DSMNRU, Lucknow UP India

pdf.yousef@lincoln.edu.my; midhun@lincoln.edu.my; ckdixit@dsmnru.ac.in

---

Abstract: Artificial Intelligence (AI) has a high potential for education, but it poses significant challenges to effective and just applications depending on crucial design choices, such as what features will be used for predication on data. This paper examines feature selection in practice. We analyze three popular real-world use cases using publicly available datasets: the UCI Student Performance dataset, the Open University Learning Analytics Dataset (OULAD), and log data from the ASSISTments intelligent tutoring platform. We explore the various features used from demographic and academic history (UCI) to behavioral Virtual Learning Environment (VLE) interactions (OULAD) and fine-grained tutoring system logs (ASSISTments). We analyze the reported impact of such features on model performance (accuracy, AUC) and evaluate the associated ethical and practical implications regarding fairness, privacy, interpretability, and actionability. Our cross-case synthesis highlights important trade-offs: granular behavioral data is often critical to achieving predictive accuracy, and yet it introduces complexities related to privacy risks [12] and is linked to challenges in interpretation, whereas the use of static or demographic features is likely to bring about direct fairness harms [4]. We conclude that careful feature selection, balancing predictive efficacy with ethical considerations, is critical for the development of AI tools that are accurate, equitable, trustworthy, and truly helpful across a range of educational contexts.

Keywords: Educational Data Mining (EDM); Student Performance Prediction; Feature Selection; Artificial Intelligence (AI); Machine Learning (ML); UCI Student Performance; OULAD; ASSISTments.

---

1. Introduction

The use of Artificial Intelligence (AI) technologies in education could have a transformative impact on personalized learning paths, early identification of students at risk of failure, and educational achievement [1]. But to responsibly unlock this potential is challenging because of bias, the need for transparency and explainability, privacy of data, and equitable access [2]-[4], [13]. Our previous work. synthesized the landscape of these opportunities and challenges, highlighting the nascent state of ethical frameworks and the need for robust evaluation [1]. This paper contributes to this discussion by considering about a core aspect of developing AI models: how to choose data features.

The selection of which features of a student one decides to include as input variables to predictive models is by no means a neutral technical task. It effectively specifies what the model can learn and has impact on its prediction accuracy and interpretability, and importantly has profound implications for fairness and equity (Baker & Hawn, 2022 [4]; Selwyn, 2019 [14]) [5]. Demographic and socio-economic background features might, for example, inadvertently capture societal biases, resulting in models that reproduce or even exacerbate already existing disparities (Baker & Hawn, 2022 [4]). Absent that connection, however, models may turn out to be not either robust or good representations of student learning experiences.

Therefore, it is important to understand the implications of feature selection in order to build reliable and effective AI in Education (AIEd) systems. This paper explores the effects of these consequences on implementations in practice by answering the following research questions:

- What categories and specific types of data features (e.g., demographic, prior academic, behavioral VLE interactions, fine-grained ITS logs) are commonly utilized in AI models predicting student performance within well-established, real-world educational datasets (UCI Student Performance, OULAD, ASSISTments)?
- According to the published research, how does the inclusion or exclusion of different feature sets appear to impact the reported predictive performance (e.g., accuracy, AUC) and the interpretability of the resulting models in these diverse contexts?
- What are apparent consequences – in terms of fairness, potential bias, data privacy, and the practical actionability of predictions - of adopting the distinct feature sets of these use cases?

For our empirical investigation, we consider the three most popular publicly available learning datasets and the related research, namely (1) the UCI Student Performance dataset, first studied by Cortez & Silva [6], (2) the Open University Learning Analytics Dataset (OULAD), established by Kuzilek et al. [7] and employed in other works such as Agudo-Peregrina et al. [8]; and from the ASSISTments system, which is commonly used in the context of Knowledge Tracing research (e.g., in Piech et al. [9]). Through the discussion of these concrete examples, we hope to shed light on the actual trade-offs involved in feature selection for AIEd.

This paper is organized as follows: Section 2 details the methodology for selecting and analyzing the use cases. Section 3 takes a deep dive into each use case, scrutinizing the features used, reported performance, and associated implications. Section 4 synthesizes findings across the cases, highlighting key patterns and trade-offs. Limitations of the study are recognized in Section 5. Finally, Section 6 concludes with a summary of findings and recommendations for future research and practice.

2. Methodology

2.1 Use Case Selection Criteria

We selected the three use cases – UCI Student Performance, OULAD, and ASSISTments – were selected based on their alignment with the following criteria, ensuring relevance, reproducibility, and diversity:

- **Real-World Origin:** Data collected from authentic educational environments (Portuguese secondary schools, UK Open University online courses, US K-12 ASSISTments platform).
- **Public Accessibility:** Datasets are available for research via established repositories (UCI ML Repository, OULAD Portal, DataShop/ASSISTments challenges), promoting transparency and further investigation.
- **Prediction Focus:** Commonly used in published research for predicting critical educational outcomes like grades, pass/fail status, dropout risk, or skill mastery.

- **Detailed Documentation:** Along with the features (variables) that are included.
- **Established Research Base:** A significant body of peer-reviewed literature utilizes these datasets, allowing for analysis of common practices and reported findings.
- **Code/Methodology Availability:** Frequent use in tutorials, open-source projects, and research papers often makes methodologies or code examples accessible (e.g., via GitHub, Kaggle).
- **Diversity of Data Types:** Collectively, they represent a meaningful spectrum of data used in AIEd:
    - UCI: Primarily static demographic, socio-economic proxies, school-related factors, and prior grades.
    - OULAD: Demographics plus rich, time-stamped logs of student interactions with a VLE.
    - ASSISTments: Highly granular, sequential log data capturing step-by-step interactions within an intelligent tutoring system..

2.2 Analysis Framework

We have looked at each case of use by combining information from key descriptive papers (e.g., [6], [7]) and representative studies performing predictive modeling tasks. The analysis framework addresses the research questions by focusing on:

1. **Contextualization:** Identifying the educational setting, the specific prediction goal(s) addressed in representative research, and the typical ML algorithms employed.
2. **Feature Identification and Categorization:** Noting the significant data features employed as predictive variables in key piece of work or stereotypical practice. Features are grouped in classes (e.g., Demographic, Prior Academic Performance, Behavioral/Interaction Logs, Contextual/School Factors, Socio-Economic Proxies, Engineered Features) to facilitate comparison.
3. **Feature Impact Analysis:** Discussion: Analysis of the reported predictiveness of individual features or groups of features based on the cited studies (e.g., feature importance scores, ablation analysis, author discussion).
4. **Performance Assessment:** Summarizing the range of predictive performance (e.g., accuracy, AUC, F1-score, RMSE) reported in relevant literature for the specific prediction tasks, noting dependencies on feature sets.
5. **Implications Evaluation:** Critically examining the consequences of using the identified feature sets, specifically addressing:
    - Fairness and Bias: Assessing the risk of models perpetuating or amplifying societal biases based on the included features (Baker & Hawn, 2022 [4]).
    - Data Privacy: Evaluating the sensitivity of the data collected and the privacy implications of its use (Klašnja-Milićević et al., 2020 [12]; Ifenthaler & Schumacher, 2021 [15]).
    - Interpretability: Considering the ease with which model predictions, based on the given features, can be understood by educators or stakeholders (e.g., Adadi & Berrada, 2018 [16]).
    - Actionability: Assessing the extent to which the features driving predictions provide timely and practical insights for pedagogical intervention..

3. Use Case Analysis

3.1 Use Case 1: UCI Student Performance Dataset

- **Context:** Prediction of final grades (G3), often divided into pass/fail, for students in secondary school mathematics and Portuguese courses in two Portuguese schools. The seminal study by Cortez & Silva [6] worked on Decision Trees, SVM, NN, and Logistic Regression. Later work commonly uses ensemble methods like Random Forests.
- **Features Used (Derived from Cortez & Silva [6]):**
  - Demographic: sex, age, address (Urban/Rural), Pstatus (Parent cohabitation status).
  - Socio-Economic Proxies: Medu (Mother's education), Fedu (Father's education), Mjob, Fjob, famsize (Family size), guardian.
  - School-Related: studytime (Weekly study time), failures (Number of past class failures), schoolsup (Extra educational support), famsup (Family educational support), paid (Extra paid classes), activities (Extra-curricular activities), nursery (Attended nursery school), higher (Wants higher education), internet (Internet access at home), absences (Number of school absences).
  - Social/Behavioral (Self-Reported): romantic (In a romantic relationship), famrel (Quality of family relationships), freetime (Free time after school), goout (Going out with friends), Dalc (Workday alcohol consumption), Walc (Weekend alcohol consumption), health (Current health status).
  - Prior Academic Performance: G1 (First period grade), G2 (Second period grade).
- **Feature Impact:** Cortez & Silva [6] proved conclusively that previous grades G1 and G2 equally dominate the final grade G3. The number of past failures also consistently ranks as highly important. Other factors showing some predictive value included higher, Medu, studytime, and schoolsup, but their influence was substantially weaker than prior academic history [6]. Several demographic and social/behavioral characteristics were found to make little contribution to predictive accuracy, particularly when G1 and G2 were used.
- **Performance:** The predictive performance as reported by Cortez & Silva [6] was strongly influenced by the number and quality of predictor variables selected. In the binary pass/fail classification, while using G1 and G2 accuracies could reach above 90%, but dropped considerably (e.g., to the 70-80% range or lower) when relying only on demographic, social, and school-related factors. Regression tasks (predicting exact G3 score) inherently yielded higher errors (e.g., Mean Absolute Error) than classification tasks. Subsequent studies with other algorithms may replicate these findings, indicating that the predictive value of previous grades is high.
- **Implications:**
  - Fairness/Bias: Excessive attention to prior (G1/G2) performance for learners can lead the models to learn just to repeat past patterns and so the "students capable of doing better" might be placed at disadvantage in front of their "students were their learning processes were not as smooth as they could have been" – it naturally detects under performers, while it does not reward the above average performances. Including SES proxies like Medu, Fedu, address, or paid directly risks embedding socio-economic bias, potentially leading to unfairly lower predictions for students from less privileged backgrounds [4].
  - Privacy: School records and questionnaires provide the data. Although less invasive than continuous behavioral monitoring, it contains sensitive self-reported information (alcohol consumption, family relationships) that must be treated with care [12].
  - Interpretability: Models like Decision Trees, as used by Cortez & Silva [6], can offer relatively interpretable rules, often dominated by simple thresholds on G1, G2, or failures.

- Actionability: Predictions heavily driven by past performance offer limited immediate pedagogical actionability beyond flagging historically low-performing students. Insights derived from potentially modifiable factors like absences or studytime (if found predictive independently) would be more actionable for teachers.

3.2 Use Case 2: Open University Learning Analytics Dataset (OULAD)

- **Context:** Predicting student outcomes (Pass, Fail, Withdrawn, Distinction) or, more commonly, identifying students at risk of withdrawal (dropout) in online courses offered by the Open University UK. Early prediction using data from the initial weeks or months is a frequent goal. Models include Logistic Regression, SVM, Random Forests, and time-aware models processing VLE interactions over time [7], [8].
- **Features Used (Based on Kuzilek et al. [7] dataset description; Agudo-Peregrina et al. [8]; common practice):**
    - Demographic: gender, region, highest_education, imd_band (Index of Multiple Deprivation - SES proxy), age_band, disability.
    - Course/Assessment: code_module, code_presentation (semester), assessment_type, score, weight, date_submitted, is_banked.
    - VLE Interaction Logs: sum_click (Total clicks across VLE resources), clicks per resource type (oucontent, url, forumng, homepage, quiz, etc.), aggregated daily or weekly click counts, number of active days, timing of first/last clicks. Engineered features capturing regularity of access, intensity, and trends over time are frequently created.
- **Feature Impact:** VLE interaction data is consistently identified as highly predictive of student persistence and success. Agudo-Peregrina et al. [8], focusing on VLE use in blended learning, found significant correlations between interaction levels (accessing resources, forum participation) and performance. In fully online contexts typical of OULAD, studies show that early and consistent VLE engagement (e.g., high sum_click in the first few weeks, regular access patterns) strongly predicts retention, while inactivity or declining activity predicts withdrawal. While demographic features like imd_band or highest_education can contribute, VLE behavioral metrics often emerge as more powerful predictors, particularly for dynamic risk assessment (e.g., Howard et al. [10]; Agustina et al., 2023 [17]). Assessment scores become highly predictive once available.
- **Performance:** To predict withdrawn students risk early based on VLE data, works using OULAD reported AUC (Area Under the Curve) values around 0.75 to 0.88, indicating the good discriminatory power. The ability to predict final pass/fail status in the course increases over the semester for the task of predicting pass/fail status with assessment information.
- **Implications:**
    - Fairness/Bias: Despite the appearance of objectivity, the use of clickstream data as a source introduces bias. Students with alternative learning preferences (e.g., offline study preference), varying access to the internet, and disabilities affecting online navigation or external time pressures (e.g., employment, caregiving) may perform less optimally on VLE activity due at least in part to reasons not related to their ability or engagement with the material per se. Application of imd_band correlates to some extent with a measure of socio-economic deprivation and would require careful ethical consideration [4].
    - Privacy: The continuous observation of minutiae of online behavior (every click, accessed resource, time spent) constitutes intensive surveillance, introducing a range of

challenges with regard to privacy—requiring clear policies, secure data management, as well as compliancy with regulations like GDPR [12]; [15].
    - Interpretability: While simple features like sum_click are interpretable, models which rely on complex temporal patterns or a large number of engineered interaction features can be opaque, and insufficiently understandable, for instructors to support understanding around any prompts for change. Developing interpretable AI (XAI) methods for such high-dimensional data are an ongoing field of research [18].
    - Actionability: High potential for timely intervention. Significant drops in VLE activity or consistently low engagement serve as clear, dynamic indicators prompting outreach from tutors or support staff. The insights are directly related to current student behavior within the learning environment.

3.3 Use Case 3: ASSISTments Datasets

- **Context:** Modeling student learning and predicting performance within the ASSISTments intelligent tutoring system, a widely-used K-12 mathematics tutor tool, and predict performance in it. Typical tasks include attempting to predict correctness on the next problem, real-time recognition of student struggle, or the monitoring and enhancement of evolving skill mastery, as in the case of Knowledge Tracing. Methodologies range from logistic regression with carefully engineered features to sophisticated sequential deep learning models like Deep Knowledge Tracing (DKT) [9] and Self-Attentive Knowledge Tracing (SAKT) [11], with recent advancements leveraging transformer architectures [19].
- **Features Used (Based on typical ASSISTments data structure and KT models):**
    - Interaction Logs: user_id, problem_id, assignment_id, skill_id/skill_name (if available), correct (Binary outcome of an attempt/problem), attempt_count, hint_count, first_action type, start_time, end_time (often used to derive time_taken).
    - Engineered Features (Common for non-sequential models): Features summarizing student history like 'number of prior attempts on this skill', 'historical success rate on this skill', 'number of hints previously used for this skill', 'time elapsed since last practice', 'counts of recent correct/incorrect responses'.
    - Sequential Features (Core for DKT/SAKT): Input is typically a sequence of student interactions represented as tuples, often (skill_id, correct) or (problem_id, correct), ordered chronologically.
- **Feature Impact:** For predicting immediate success, features like correct on the immediately preceding attempt(s), hint_count, and attempt_count on the current problem are strong predictors. For Knowledge Tracing models such as DKT, relationship between interactions over time (in particular correctness) with respect to certain skills is the key information that enables the model to estimate latent knowledge states [9]. The model also tacitly accounts for the concepts of learning decay (forgetting), as well as the influence of practice. These engineered features representing student history (e.g. success rates, counts) are vital to improving the performance of non-sequential predictive models on this data.
- **Performance:** Deep Knowledge Tracing (DKT) applied to large ASSISTments datasets demonstrated significant improvements over traditional methods like Bayesian Knowledge Tracing, achieving AUC values for predicting next-problem correctness often in the range of 0.70 to 0.85, depending on the specific dataset split and evaluation setup [9]. Newer models like SAKT [11] and transformer-based approaches often report further improvements [19]. Models predicting simpler outcomes like immediate attempt correctness can reach higher accuracy.
- **Implications:**

- Fairness/Bias: Models might pick up on biases as a part of user interaction patterns. For instance, the model might perceive that certain groups of students are faster to ask for a hint or to try multiple, rapid responses (‘gaming the system’), the model might interpret this differently regarding underlying knowledge compared to students who work slowly and deliberately. Inaccuracies or biases in the underlying skill tagging within the platform could also lead to biased knowledge estimates [4].
- Privacy: Involves extremely fine-grained logging of every student interaction within the platform. Given the K-12 context, this necessitates stringent anonymization, data security, and ethical oversight concerning data usage and parental consent [12].
- Interpretability: Deep learning models like DKT and SAKT are accused of being “black boxes”. Although models can have high predictive accuracy, their internal representations of student knowledge may be difficult for human teachers to understand, which could result in limited insight into why a model predicts a student to master or fail with a certain skill. While the Explainable AI for Knowledge Tracing (XKT) is an open research challenge and has been tried to be addressed with different approaches [20], [21]
- Actionability: High potential for automated, within-platform actionability. Predictions about skill mastery or struggle can directly inform the system's decisions about providing hints, selecting the next problem (adaptive sequencing), or offering specific feedback, creating a responsive learning environment. Translating these fine-grained insights into actionable information for teachers outside the platform requires effective summarization and reporting.

4. Cross-Case Synthesis and Discussion

Comparing these three examples shows important patterns about choosing data features in AI for education (AIEd):

- **Different Data, Different Strengths (Data Spectrum and Predictive Focus):** The examples show a range from static, background-rich data (UCI) to active, combined behavioral data (OULAD) to very detailed, step-by-step interaction data (ASSISTments). The type of data naturally shapes what the AI tries to predict. UCI data (with early grades like G1/G2) is good at predicting future grades based on past ones [6]. OULAD's online activity logs are good for predicting dropout risk based on student engagement [8]. ASSISTments' detailed logs help in modeling how students learn new skills step-by-step [9].
- **Recent Performance Matters Most (Feature Importance Hierarchy):** A common pattern is that what a student did recently is often a very strong sign of what they will do next. Prior grades strongly predict future grades (UCI). Current online activity strongly predicts if a student will stay in a course (OULAD). And recent answers in software strongly predict how they'll do on the next problem (ASSISTments). While background information (like family or school details) can add some information, it's often less powerful than recent academic or behavioral data. However, using background data still has important ethical considerations (Baker & Hawn, 2022 [4]).
- **Balancing Accuracy with Other Concerns (Performance vs. Ethical/Practical Trade-offs):** There's often a tough choice between making AI predictions as accurate as possible and dealing with ethical and practical problems:
  - Accuracy vs. Fairness: Data features that give high accuracy (like past grades, or measures of family income like imd_band or Medu) might also carry the biggest risk of being unfair or repeating old biases. Just using data about what students do (clicks,

attempts) isn't always fair either, as how students interact online can also be affected by outside inequalities (Baker & Hawn, 2022 [4]).

  - Accuracy vs. Privacy: Using rich data about student behavior and interactions (OULAD, ASSISTments) usually leads to more accurate, up-to-date predictions. But it means collecting a lot of detailed data, which increases worries about student privacy (Klašnja-Milićević et al., 2020 [12]).
  - Accuracy vs. Understanding: Complex AI models that use complicated data (like the step-by-step data in ASSISTments) often get the most accurate results. But they can be harder to understand compared to simpler AI models that use more straightforward data (like Decision Trees with the UCI data). This highlights the need for robust XAI methods tailored to education (Adadi & Berrada, 2018 [16]; Ud-Din & Lee, 2022 [20]).
- **How Helpful is the Prediction? (Contextual Actionability):** How useful a prediction is for teachers depends a lot on the data features used. Data about current student actions (like online clicks in OULAD or software interactions in ASSISTments) gives timely signals that can lead to quick help (either from a person or automatically from the software). Predictions based on static or old data (like student background or past Fs in UCI) are less useful for immediate teaching changes; they often just point out students who were already known to be at risk.

To summarize these comparisons visually:

Table 1. Comparison of Use Case Characteristics

Feature	UCI Student Performance	OULAD	ASSISTments Datasets
Primary Data Type	Static Records & Surveys	VLE Clickstream Logs	ITS Interaction Logs
Key Predictive Features	Prior Grades (G1, G2), Past Failures	VLE Click Count, Activity Timing, Scores	Interaction Sequence (Skill, Correctness)
Common Goal	Predict Final Grade/Pass/Fail	Predict Dropout/Course Outcome	Predict Next Answer/Knowledge State
Typical Performance	Accuracy: High (with G1/G2)	AUC: Good-High (for dropout)	AUC: Good-High (for next problem)
Main Fairness Risk	Using SES Proxies, Relying on Past Results	Unequal Access/Use Patterns, SES Proxy	Interaction Style Bias, Skill Definition
Main Privacy Concern	Sensitive Survey Data	Detailed Online Activity Tracking	Granular Step-by-Step Tracking
Interpretability	Potentially High (Simpler Models)	Moderate (Click counts) to Low (Patterns)	Often Low (Deep Learning Models)
Actionability	Lower (Based on History)	High (Timely Engagement Signal)	High (Immediate In-System Feedback)

This table presents the key differences and similarities between the three data examples in an easily understandable format, and adds evidence for the discussion made above. It further demonstrates that there is no single “best” type of data; the answer depends on the prediction target, the available resources, and critically, the trade-off between ethical considerations in using different types of features.

5. Limitations

This study, by focusing on some examples, has its limitations. It is an aggregation of results from several existing research, and it really depends on the level of detail on feature engineering and importance per publications. Quantitative comparison quickly across different data sets, prediction tasks and populations is hard. Also, this experiment is mostly about feature selection, in addition to the selection of the ML algorithm and the hyperparameter tuning also heavily affect the result. Third, it needs to be cautioned that drawing causal inference between individual points and results within the predictive modeling is not appropriate

6. Conclusion and Future Directions

This comparison of three leading AIEd use cases makes it clear that feature selection is not a mere technical preliminary but a critical decision point laden with ethical and practical consequences. The journey from static student attributes (UCI) through VLE interactions (OULAD) to granular ITS logs (ASSISTments) reveals an evolving landscape where predictive power often increases with data granularity but concurrently raises challenges related to fairness, privacy, and interpretability. Previous academic performance continues to be a strong predictor [6], but uncritical use is adopted at the risk of strengthening existing inequities [4]. Behavioral data offers dynamic insights [8] but demands careful ethical stewardship [12], [14]. Fine-grained interaction information allows complex modeling [9], [11], [19]often by using more recent architectures, but lacks interpretability.

Developing responsible and effective AIEd necessitates a holistic approach to feature selection, explicitly weighing predictive goals against potential biases, privacy intrusions, and the need for understandable and actionable insights. Moving forward, the field requires concentrated effort in several key areas:

1. **Rigorous Comparative Studies:** Perform studies that compare the models built using various combinations of features (static, behavioural, hybrid) on the same dataset designed for such a purpose, not just using accuracy but also fairness metrics (i.e., subgroup parity) and interpretability.
2. **Proactive Bias Auditing and Mitigation:** The incorporation of fairness audits and bias mitigation approaches as a natural part of model construction focusing on the influence of widely used education features [4].
3. **Development of Meaningful, Privacy-Conscious Features:** Investigating and testing features that can best capture pedagogical constructs (e.g., learning strategies, self-regulation) in a meaningful way that is less dependent the over-sensitive or highly-biased raw data [12]
4. **Advancing Explainable AI (XAI) in Education:** Tailoring and developing XAI methods for educational settings so that educators can comprehend and trust the predictions of complicated models or features [20], [18].
5. **Closing the Loop - Evaluating Intervention Impact:** Focusing research on the real-world effectiveness and equity implications of interventions triggered by AI predictions based on different feature sets, moving beyond purely predictive performance evaluation.

By taking a more critical, ethically informed, and contextually responsive approach to which data we use to fuel our AI models, we may have a better chance of ensuring that these powerful technologies have positive and equitable impact on the future of learning.

References

1. M. Y. AlFaress, M. Janarthanan, and C. K. Dixit, "Predicting Student Performance and Enhancing Equity with AI: A Literature Review," SGS Engineering & Sciences, vol. 1, no. 1, LGPR, 2025, https://spast.org/techrep/article/view/5278.
2. S. Akgun and C. Greenhow, "Artificial intelligence in education: Addressing ethical challenges in K-12 settings," *AI and Ethics*, vol. 2, no. 3, pp. 431-440, 2022. DOI: 10.1007/s43681-021-00096-7
3. J. Borenstein and A. Howard, "Emerging challenges in AI and the need for AI ethics education," *AI and Ethics*, vol. 1, no. 1, pp. 61-65, 2021. DOI: 10.1007/s43681-020-00002-7
4. Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education, 32*(4), 1052-1092. DOI: 10.1007/s40593-021-00285-9
5. R. A. Poldrack, G. Huckins, and G. Varoquaux, "Establishment of best practices for evidence for prediction: A review," *JAMA Psychiatry*, vol. 77, no. 5, pp. 534–540, 2020. **DOI: 10.1001/jamapsychiatry.2019.3671**
6. P. Cortez and A. Silva, "Using Data Mining To Predict Secondary School Student Performance," in *Proc. 5th Future Business Technology Conf. (FUBUTEC 2008)*, 2008, pp. 5-12. Available at: University of Minho Repository
7. J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics dataset," *Scientific data*, vol. 4, no. 1, Art. no. 170171, 2017. DOI: 10.1038/sdata.2017.171
8. Á. F. Agudo-Peregrina, S. Iglesias-Pradas, M. Á. Conde-González, and Á. Hernández-García, "Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F learning," *Computers in Human Behavior*, vol. 31, pp. 542-550, Feb. 2014. **DOI: 10.1016/j.chb.2013.06.031**
9. C. Piech et al., "Deep knowledge tracing," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 28, 2015. Available at: arXiv:1506.05908
10. E. Howard, M. Meehan, and A. Parnell, "Contrasting prediction methods for early warning systems at undergraduate level," *The Internet and Higher Education*, vol. 37, pp. 68-75, Apr. 2018. DOI: 10.1016/j.iheduc.2018.02.001
11. S. Pandey and G. Karypis, "A Self-Attentive model for Knowledge Tracing," in *Proc. 12th International Conference on Educational Data Mining (EDM)*, 2019. DOI: 10.48550/arXiv.1907.06837
12. Klašnja-Milićević, A., Ivanović, M., & Nanopoulos, A. (2020). Privacy-preserving educational data mining: A literature review. *Education and Information Technologies, 25*, 1139-1162. **DOI: 10.1007/s10639-019-10032-2**
13. Holstein, K., & Doroudi, S. (2021). Equity and artificial intelligence in education: Will “AIEd” result in equitable and just learning opportunities?. In *AI and education* (pp. 23-34). Springer, Cham.
14. Selwyn, N. (2019). What’s the problem with learning analytics?. *Journal of Learning Analytics, 6*(3), 11-19. DOI: 10.18608/jla.2019.63.3

15. Ifenthaler, D., & Schumacher, C. (2021). Student perceptions of data privacy in learning analytics. *Educational Technology Research and Development, 69*(5), 2661-2679. DOI: 10.1007/s11423-021-10043-2

16. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE access, 6*, 52138-52160. DOI: 10.1109/ACCESS.2018.2870052

17. Agustina, R., Lee, C. H., & Liu, C. C. (2023). Enhancing dropout prediction in online learning using interpretable machine learning and engagement features. *Computers & Education, 204*, 104866.

18. Van der Waa, J., Van Diggelen, J., Van Den Bosch, K., & Neerincx, M. (2021). The explainable AI life cycle. *Frontiers in Artificial Intelligence, 4*, 628100.

19. Abdelrahman, G., & Wang, Q. (2023). A review of transformer-based knowledge tracing models. *Education and Information Technologies*, 1-31.

20. Ud-Din, S., & Lee, Y. K. (2022). A survey on explainable knowledge tracing. *IEEE Access, 10*, 90509-90527.

21. Guo, L., Zhang, Z., Lin, X., & Yu, Z. (2021). Interpretable knowledge tracing: A survey. *arXiv preprint arXiv:2112.01384*