Predicting Student Performance: Advocating Transparent Models Over Black Box Approaches

Aims & Objectives:

- • Investigate traditional explainable machine learning techniques, including Explainable Boosting Machines (EBMs), decision trees, and rule-based models, to evaluate their effectiveness in predicting student success in CS1
- • Compare these techniques with each other and the (Predict Student Success) PreSS model to find the best balance between interpretability and accuracy
- • Create a better prediction model using clear features that provides actionable insights for educators.

Research Methods:

- • PreSS Model (Predict Student Success)
- • EBMs (Explainable Boosting Machines)
- • Decision Trees
- • Rule-Based Models

We will evaluate and compare various ML techniques to identify the most effective and accurate ones for predicting educational outcomes in CS1 students.

Abstract & Introduction:

Predicting a student's success in introductory programming courses (CS1) is essential for identifying at-risk learners and promoting actions to improve educational results. While the Predict Student Success (PreSS) model has been previously employed to forecast student performance using various factors, more complex methods like deep learning have not significantly improved accuracy or transparency as expected. There have been many criticisms regarding the use of explaining complex black box models. Instead, it is beneficial to use inherently interpretable machine learning (ML) models, which offer clear and accurate explanations based on specific domain knowledge.

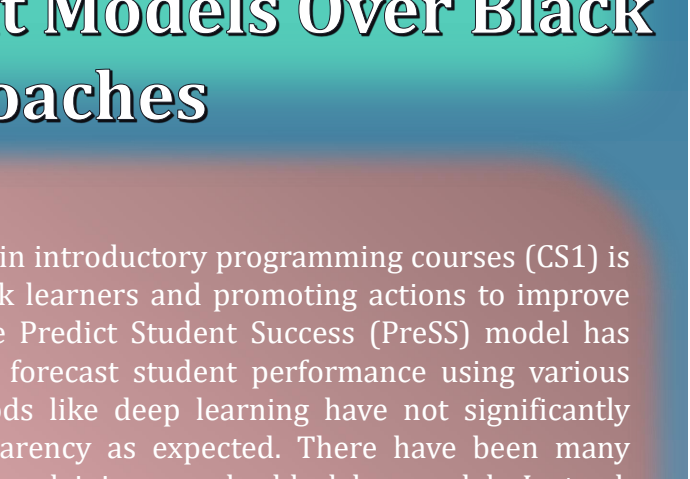
Many machine learning systems operate as black boxes, resulting in serious consequences due to their opacity. We aim to explore the challenges posed by these black box systems and the techniques for explainable machine learning (ML), highlighting the critical need for model transparency in high-stakes decision-making. Interpretability is important in educational settings; this thesis proposes a systematic examination of traditional, explainable machine learning techniques, such as Explainable Boosting Machines (EBMs), decision trees, and rule-based models, in predicting CS1 outcomes.
This research focuses on gaining clarity and understandable methods to assist educators by gaining insights into what drives student success and failure.

Results and Findings:

Black Box Models

- Frequently resulting in reduced interpretability and uncertain gains in prediction accuracy
- It can be challenging to criticize the decision-making process or identify any shortcomings in the conclusions dra
- Many users find them less understandable and trustworthy than interpretable machine learning techniques

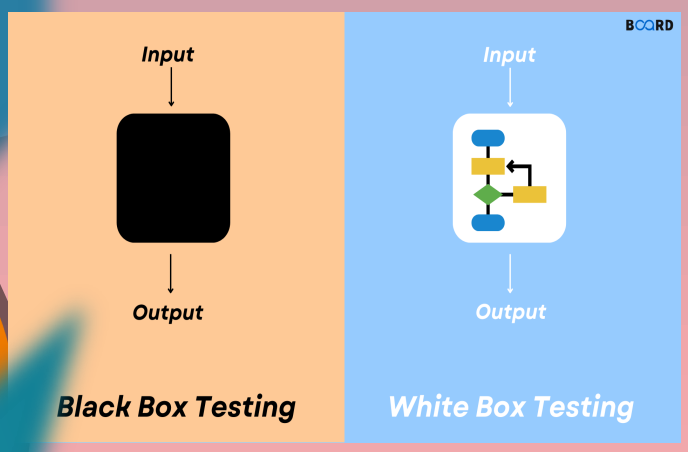
Examples → Deep Learning and Large Language Models (LLMs)



Interpretable Machine Learning Techniques

- Are "inherently interpretable" meaning they are easily understood without further explanation required
- Trust is enhanced by making model decisions understandable to users
- Accuracy and performance are often improved due to trust and detecting bias

Examples → (Explainable Boosting Machines) EBMs, decision trees, and rule-based models



Conclusion:

In conclusion, we must challenge the prevailing notion that black box models are essential for achieving accurate predictions in Explainable Machine Learning. We need to encourage policymakers to prioritize interpretable models over black box ones such as deep learning and LLMs, and to be aware of the current challenges in interpretability. The study stresses the urgency of preventing black box models from being accepted without adequate scrutiny, as their use can lead to significant societal risks and poor decision-making in critical areas such as criminal justice, public safety, healthcare, and the education system.
A proposed direction for future research could be the implementation of robust Interpretable Models. By leveraging research, we can develop new algorithms and methodologies that inherently emphasize interpretability, allowing for more trust in machine learning systems in high-stakes scenarios.
it is essential to establish guidelines and standards for evaluating the interpretability of machine learning models. Holding organizations accountable for their use of black box systems is particularly vital in sensitive applications.

Image References:

[Image 1] "White-box Machine Learning: How the Model Works & Top Benefits," Unit21.ai, 2025. https://www.unit21.ai/fraud-aml-dictionary/white-box-machine-learning (accessed Jan. 30, 2025).
[Image 2] "Black Box vs White Box: Software Testing," Board Infinity, Apr. 08, 2023. https://www.boardinfinity.com/blog/white-box-vs-black-box/ (accessed Jan. 30, 2025).

References:

[1] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1, 206–215 (2019). https://doi.org/10.1038/s42256-019-0048-x (accessed Jan. 30, 2025).
[2] Quille, K., & Bergin, S. (2019). CS1: how will they do? How can we help? A decade of research and practice. Computer Science Education, 29(2–3), 254–282. https://doi.org/10.1080/08993408.2019.1612679 (accessed Jan. 30, 2025).
[3] Keith Quille and Susan Bergin. 2018. Programming: predicting student success early in CS1. a re-validation and replication study. In Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE 2018). Association for Computing Machinery, New York, NY, USA, 15–20. https://doi.org/10.1145/3197091.3197101 (accessed Jan. 30, 2025).
[4] Keith Quille, Lidia Vidal-Meliá, Keith Nolan, and Aidan Mooney. 2023. Evolving Towards a Trustworthy AIEd Model to Predict at Risk Students in Introductory Programming Courses. In Proceedings of the 2023 Conference on Human Centered Artificial Intelligence: Education and Practice (HCAIep '23). Association for Computing Machinery, New York, NY, USA, 22–28. https://doi.org/10.1145/3633083.3633190 (accessed Jan. 30, 2025).
[5] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nature Machine Intelligence, vol. 1, no. 5, pp. 206–215, May 2019, doi: https://doi.org/10.1038/s42256-019-0048-x (accessed Jan. 30, 2025).
[6] E. giosmin, "Interpretability vs explainability: Understanding the Differences and Importance in the World of Artificial Intelligence," XCALLY Motion, Jul. 04, 2023. https://www.xcally.com/news/interpretability-vs-explainability-understanding-the-importance-in-artificial-intelligence/ (accessed Jan. 30, 2025).

Acknowledgements:

The HCAIM (the Human-Centred AI Master's Programme) Project is Co-Financed by the Connecting Europe Facility of the European Union Under Grant №CEF-TC-2020-1 Digital Skills 2020-EU-IA-0068. This poster was created as part of the Blended Intensive Programme organised under the Erasmus + Programme of the European Union.

By Olohigbe Pearl Ohiwerei