Using Machine Learning Techniques to Solve Problems in Materials Science & Engineering

Eryn Dennis, Simon Billinge, Ph.D., College of Information Sciences & Technology, Pennsylvania State University

Introduction: As STEM fields become increasingly interdisciplinary, fields that didn't previously require programming knowledge can now use machine learning to solve problems. Educational examples (edexes) are hands-on exercises that illustrate the use of machine learning to solve scientific problems. Edexes can be used across several STEM disciplines, however, the focus for the project is undergraduate and master's students studying Materials Science & Engineering (MATSE). Used as ipynbs, edexes serve as a space for students to practice and apply their knowledge of a topic using machine learning and are stored in a database for professors to access at any time. Our project consists of a software infrastructure, which is a setup to make the database and the edexes simple for a professor to use. This infrastructure will allow a professor to use edexes to build a course to their liking. To complete the infrastructure, we created edex course builder (ecb) tools that the professors will be interacting with to create their courses. Some of these tools include a filtering tool, a labeling tool, and a sequencing tool. The eventual goal for this project is for it to expand and become an openly shared community website that professors worldwide can use to select and sequence edexes for their courses and contribute their own edexes.

Methods: To make our software infrastructure straightforward for professors to interact with, we created a process for selecting, sequencing, and uploading edexes. We developed edex course builder (ecb) tools, which the professors will use throughout the process. Once an edex is done being created, it gets placed in the database and is labeled by the creator based on its content. Once a professor wants to choose edexes for their course, the edexes are passed through our first ecb tool, which is a filtering tool. Once filtered, the remaining edexes are passed through the next ecb tool, the sequencing tool, and returned to the professor. These two steps can be repeated until the professor is satisfied with the list of edexes given. Once satisfied, the professor chooses which edexes they want to include in their course from the suggested edexes in the list. The chosen edexes are confirmed by the professor and are then uploaded to Canvas in the appropriate modules based on their content.

Results: We were able to design a process for edex selection, sequencing, and uploading that's simple for professors to use. We also produced a JSON file to store information about each edex, which is where each edex will be labeled once added to the database. We also successfully created a filtering function and started producing our own edexes.

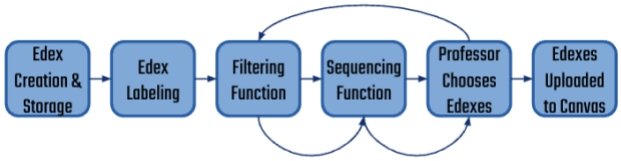

Figure 1. Our process for selecting, sequencing, and uploading edexes. This process ensures that professors can easily decide which edexes to use for their courses. Refer to the Methods section for more information.

Conclusions: Edexes prove to be a great way for MATSE students to practice machine learning techniques. The software infrastructure and ecb tools were designed and developed to make course creation faster and more efficient for professors. This project also allows for easy collaboration with other professors, departments, and universities. The software was built in a way where it can be easily expanded to a website that can be openly used by a community of professors around the world. Therefore, this project can be beneficial for both students and professors globally.

Acknowledgements: I would like to thank the SURE program for giving me the opportunity to do this research, Professor Billinge for allowing me to work on this project and in his lab, and all the members of the Billinge Group for mentoring me and helping me navigate the lab.