Refactoring python classes

Bryan Aleman, Elgin Community College, Elgin, IL 60123
Robert Schaffer, Jennefer Maldonado, National Synchrotron Light Source II, Brookhaven National Laboratory, Upton, NY 11973

The National Synchrotron Light Source II (NSLS-II) at Brookhaven National Laboratory is a U.S. Department of Energy (DOE) user facility that provides extremely bright X-ray light for scientific research. Each experimental station, known as a beamline, uses shared software tools to control motors, detectors, and other devices during experiments. Over time, separate beamlines have independently developed similar or identical software components, creating unnecessary code duplication. This project investigated that duplication with the goal of improving software maintainability and supporting more efficient scientific operations at NSLS-II. Working within the Data Science and System Integration (DSSI) and Data Acquisition and Detectors (DAD) group at NSLS-II, I analyzed the Python software libraries used to operate beamline hardware. I developed automated methods to extract, compare, and classify thousands of device-control classes from multiple beamlines. The analysis first used an AI-based similarity clustering approach and then a deterministic Python method that produced inheritance maps, attribute/method inventories, and pairwise similarity scores using a Jaccard metric. A total of 1,978 classes were cataloged and multiple sets of similar device classes were identified across different beamlines. These results show that consolidation of repeated code is both necessary and feasible. Reducing duplication can shorten development time for future beamlines and make the NSLS-II control software more reliable, which supports the DOE mission of enabling high-impact scientific research with the use of world-class facilities. Through this project, I gained experience working in a large-scale scientific computing environment, developed automation tools for code analysis, collaborated with engineers, and learned best practices for sustainable scientific software development.