The US Department of Energy publishes Incident Report documents to record events that go wrong in past projects to prevent them from happening again. Argonne National Laboratory, our partner organization, realized they were not making use of these reports when beginning new projects due to the overwhelming volume of available data. To help identify when a project might benefit from lessons learned from past projects, Argonne created a machine learning workflow to match both existing and new Incident Reports with Project Planning Documents. This notifies the appropriate personnel of relevant incidents and prevents specific mistakes from being repeated.

The goal of the project was to adapt the machine learning workflow used by Argonne to suit similar-style documents at Fermilab. Argonne and Fermi have different data formatting and management systems, so we had to adapt the pipeline accordingly. To ameliorate these differences, we cleaned and preprocessed the raw Fermi Project Planning Documents by removing formatting errors and joining multiple rows of the Fermi data by their unique Project Planning Document ID. Next, we adjusted the code to have the option to switch various features on or off: for example, when running the code, there are options for web-scraping and whether to send an email notification to the appropriate project planning personnel. These flags make it easier for Fermi to test out the different features to decide what would be relevant for their internal use.

We found that the pipeline works well on the Fermi data. By comparing the unique tokens in the Incident Reports and recommended Project Planning Documents, we found the matches to be qualitatively relevant. We believe that our modifications will ultimately help to enhance workplace safety by minimizing repeat errors at Fermi.

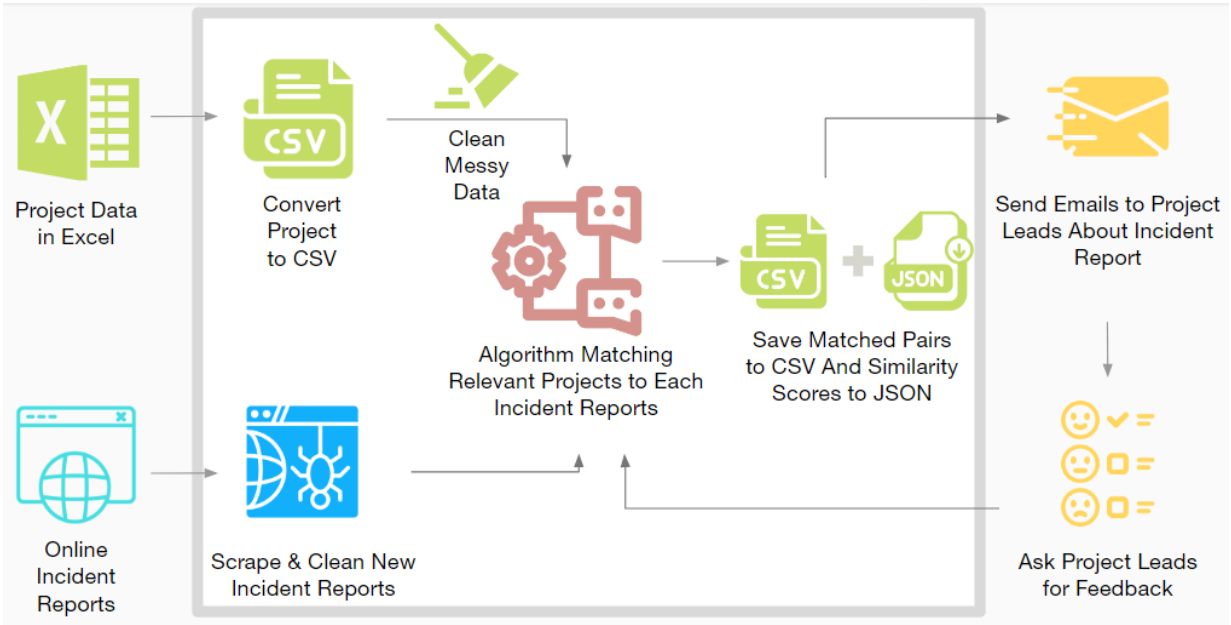

Figure 1: Fermi Data Pipeline (Our Contribution in Gray Box)