

Week 3: Team Log

This week we made improvements to our sprint workflow by creating smaller tasks in Jira. This was to make the tasks more granular so they could be completed within our weekly sprints. We also made it a goal to update Jira more often so that people can easily see the progress of each task without having to ask people directly. One major step from this week is that we identified inconsistencies within the data, which was found to almost all be within the 'units' column. After identifying the inconsistencies, we created code to be able to handle them. We also looked into what tools we can use for data flow connections. At this point we are mainly focused on creating a python script for clustering and classifying the tags but will need to focus on data flow connections once we finish that. We also identified NRCan secondary end-use classifications but after further investigation we have decided to change our approach to labelling and clustering. This is because after our meeting with UDL, we discovered that the NRCan tags are only for energy consumption items. This means we are changing our approach to how we are labelling our training data. Overall, we are on track for modelling and testing for this week and we feel as a team we have been working together cohesively.

Tasks for the Next Week

1. Create testing and training data
2. Feature engineering
3. Implement feature selection techniques
4. Research semi-supervised methods