

Weekly Sprint Planning

2020-06-05 / 10:00-11:30 / Zoom

	WEEKS	DATES	GOALS
Investigation and Research	3	27 April - 14 May	Identify project objectives and key data features + understand data dictionaries + research machine learning techniques to classify building sensors
Training/Testing Data and Data Prep	1	15 - 21 May	Create training/testing data + transform data for machine learning tasks
Feature Selection and Engineering	1	22 - 28 May	Aggregate data + create smaller categorical levels + identify relevant features
Initial Modeling	1	29 May - 4 June	Develop a classification model to apply group tags to end-uses for the Pharmacy building + finalize main script to clean data and feed data into models + UBC mid-term presentation
Tuning Model	1	5 - 11 June	Validate and evaluate models
Dashboard and Wrap-Up	1	12 - 18 June	Create visualizations of results + data pipeline of results + complete user-acceptance testing of dashboards + start final report + unit testing
Wrap-Up	1	19 - 26 June	Final report + package final code + complete unit testing + UDL final presentation + UBCO final presentation
Total Weeks	9		

1. What was our goal/theme from last week?

Goal: Initial modeling

2. Which tasks did we complete?

- Finished `append_agg` function
- Decided we want lasso, but changed to ridge
- Code to calculate sensor update
- Code to aggregate string values
- Code to aggregate boolean values
- Code to aggregate numeric values
- Completed feature selection
- Identified relevant features

- Completed code to scale values
 - Implemented feature engineering
3. Was there anything stopping us from finishing specific tasks?
 - Not enough time
 - Waiting on Arthur's diagrams
 4. What tasks are still in progress?
 - Making training and testing data
 - Develop code to grid search
 - Determine when Skyspark stop making changes
 - Model EC and NC relationship
 - Populating main function
 5. Are there any changes that need to be made?
 - a. No
 6. What is our goal/theme for this week?
Goal: Tuning model
 7. What tasks need to be added/replenished to the Backlog?
 - a.
 8. What tasks are most important and should be pulled from Backlog to In progress?
 - Task 1
 - Task 2
 - Task 3
 9. Are there any dependencies between In Progress tasks?
 - a. If so, how will that be organized?
 10. Who is going to be assigned to which tasks and update in Jira?

Person	In progress Tasks	New Tasks
Claudia	•	•
Connor	•	•
Eva	•	•
Alex	•	•

- **Populating main.py**
 - Step 1 - Cluster NC Data (Connor)
 - Step 2 - Model EC/NC Relationship (Eva)
 - Step 3 - Mid-Process Cleanup (Alex)
 - Step 4 - Prep EC Data for Classification Model (Claudia)
 - Step 5 - Classification Model for EC Data (Connor)
 - Step 6 - Output results to InfluxDB (Alex)
- **Tuning the model - Increasing accuracy**
 - Research ways to evaluate models/choose performance metrics for comparison (Connor)
 - Develop function for the chosen performance metrics (if needed) (Connor)
 - Outline reasons for chosen models (Eva)
 - Develop code to grid search for optimal model/model configuration/features (tentative)
 - Sklearn has something for this that works well would just need to code it up
 - Recommend models
 - Create a flowchart/report to show chosen model and outputs (Claudia)
- **Optimize and clean the code - Increasing speed**
- **Develop various classification models for NRCan tags (Alex and Connor)**
 - Test with our Step 4 data (feed our data into Connor's awesome function) (all)

- Make the function with chosen model (Connor for now, and all to optimize)