Weekly Sprint Planning

2020-05-15 /10:00-11:30 / Zoom

THEMES	WEEKS	DATES	GOALS
Investigation and Data Prep	3	27 April - 14 May	Identify project objectives and key data features + understand data dictionaries + transform data for machine learning tasks.
Model	1	15 - 21 May	Develop a classification model to apply group tags to end-uses for the Pharmacy building.
Model	1	22 - 28 May	Validate and evaluate models.
Scale + Analysis	2	29 May - 11 June	Expand the model to other UBC buildings (if time permits) + complete user-acceptance testing of model + identify conclusions + create visualizations of results + complete user-acceptance testing of dashboards + UBC mid-term presentation
Wrap-Up	2	12 - 26 June	Final report + package final code + UDL final presentation + UBCO final presentation
Total Weeks	9		

1. What was our goal/theme from last week?
   Goal: Investigation and Data Prep

2. Which tasks did we complete?
   - **Confirmed NRCan Classifications**
   - **1-Page Summary of Group Progress**
   - **Submitted Proposal**
   - **Review sensor metadata**

3. Was there anything stopping us from finishing specific tasks?
   - Haven't heard back from Jiachen about the building data meeting he had with Arthur
   - Delayed confirmation of NRCan classifications

4. What tasks are still in progress?
   - Transform Data for ML Tasks
   - Identify Relevant Features

- Confirm Details on Project Objectives
- Develop a tool to identify and populate missing/inconsistent data

5. Are there any changes that need to be made?
   - Update Jira with comments more regularly

6. What is our goal/theme for this week?
   Goal: Feature selection and model testing

7. What tasks need to be added/replenished to the <u>Backlog</u>?
   - Make training and testing data
   -

8. What tasks are most important and should be pulled from <u>Backlog</u> to <u>In progress</u>?
   - Task 1
   - Task 2
   - Task 3

9. Are there any dependencies between <u>In Progress</u> tasks?
   a. Labelling data (lots of tasks depend on this)

10. Who is going to be assigned to which tasks and update in Jira?

- **Tuesday's presentation (Claudia)**
  - Create presentation powerpoint
- **Make training and testing data**
  - Manually label the data (Claudia & Alex)
  - Clarify with UDL regarding dual-classifications / research more about the correct classification
  - Review the labels (Eva and Connor)
- **Feature engineering**

- ○ Identify if any feature engineering can be done to aid in classification, possible examples include the following:
    - ■ Extracting substrings from equipment/sensor naming schemes
    - ■ Variability of measurement values by a given piece of equipment or sensor dependent on environmental conditions
    - ■ Update-rate of sensor (there are some that update ~1 per day, others that update at set intervals, others that vary greatly)
    - ■ Aggregating measurements over a set time (ex. Weekly average temperature)
- **Feature selection**
    - ○ Join metadata and SkySpark data (all data, pharmacy only?) (Eva)
    - ○ Identify relevant metadata fields (Alex)
        - ■ Confirm with Jiachen regarding metadata field selections
    - ○ Decide training/testing split - 70/30?
    - ○ Decide which feature selection techniques to implement
    - ○ Implement feature selection technique 1
    - ○ Implement feature selection technique 2
    - ○ Compare the results to decide which features
- **Research semi-supervised models** (Connor)
- **Develop various models for classification**
    - ○ Decide which models to implement (meet and brainstorm?)
    - ○ Implement model 1
    - ○ Implement model 2
    - ○ Implement model 3
    - ○ Implement model 4