Status Report:
UBCO MDS Capstone - Urban Data Labs

WEEK 6

Connor Lee, Claudia Nikel, Eva Nguyen, Alex Tamm

Outline

- Progress made during previous week
  - Individual logs
  - Team logs
- Current Progress
- Preliminary Results
- Difficulties & Roadblocks
- Plan for next cycle

Previous Week’s Progress

Progress - Individual Work Logs

Connor

Created function that combines aggregations together + integrated clustering into main function + researched packages for supervised classification + created template for comparing model performance + help with presentation

Claudia

Research model for EC/NC relationship + create code for Ridge Regression model + create midterm presentation + reviewed code in main.py file + coded for grid search to find optimal alpha value for regression model

Alex

Looked into more details of database update + created pseudo code file with inputs/outputs for each step in model + worked on finding ideal dataset date range + finished uom code update + help with presentation

Eva

Research model for EC/NC relationship + prepped data for Ridge Regression code + created code for Ridge Regression model + created dummy dataset to feed into classification model + help with presentation

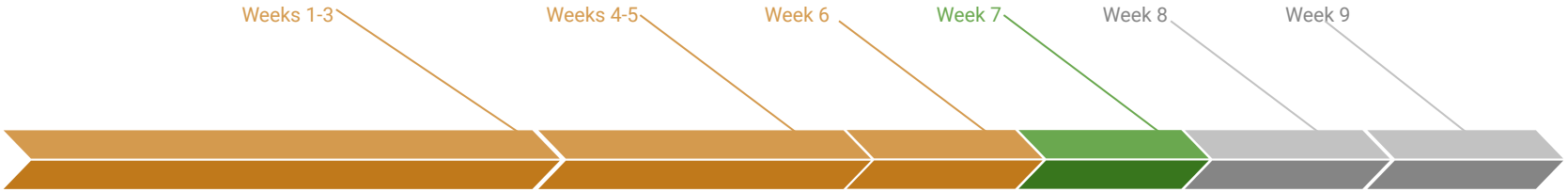
Progress - Team Work Logs

Accomplishments

- Midterm presentation
- Output from clustering model
- Output from regression model
- Finished aggregation functions
- Finished feature selection & identified relevant features
- Created code for scaling values
- Implemented feature engineering

Current Progress

Project Schedule



Investigation & Data Prep

- Identify project objectives and key data features
- Understand data dictionaries
- Transform data for machine learning tasks

Feature Selection/Engineering

- Research feature selection techniques
- Merge data & metadata
- Make categorical data into smaller fields
- Aggregate different values
- Identify relevant continuous & categorical features
- Create testing and training data

Initial Modelling

- Create 3 models for each step in our project
- Run test through main.py with test dataset to get a result

Model Tuning

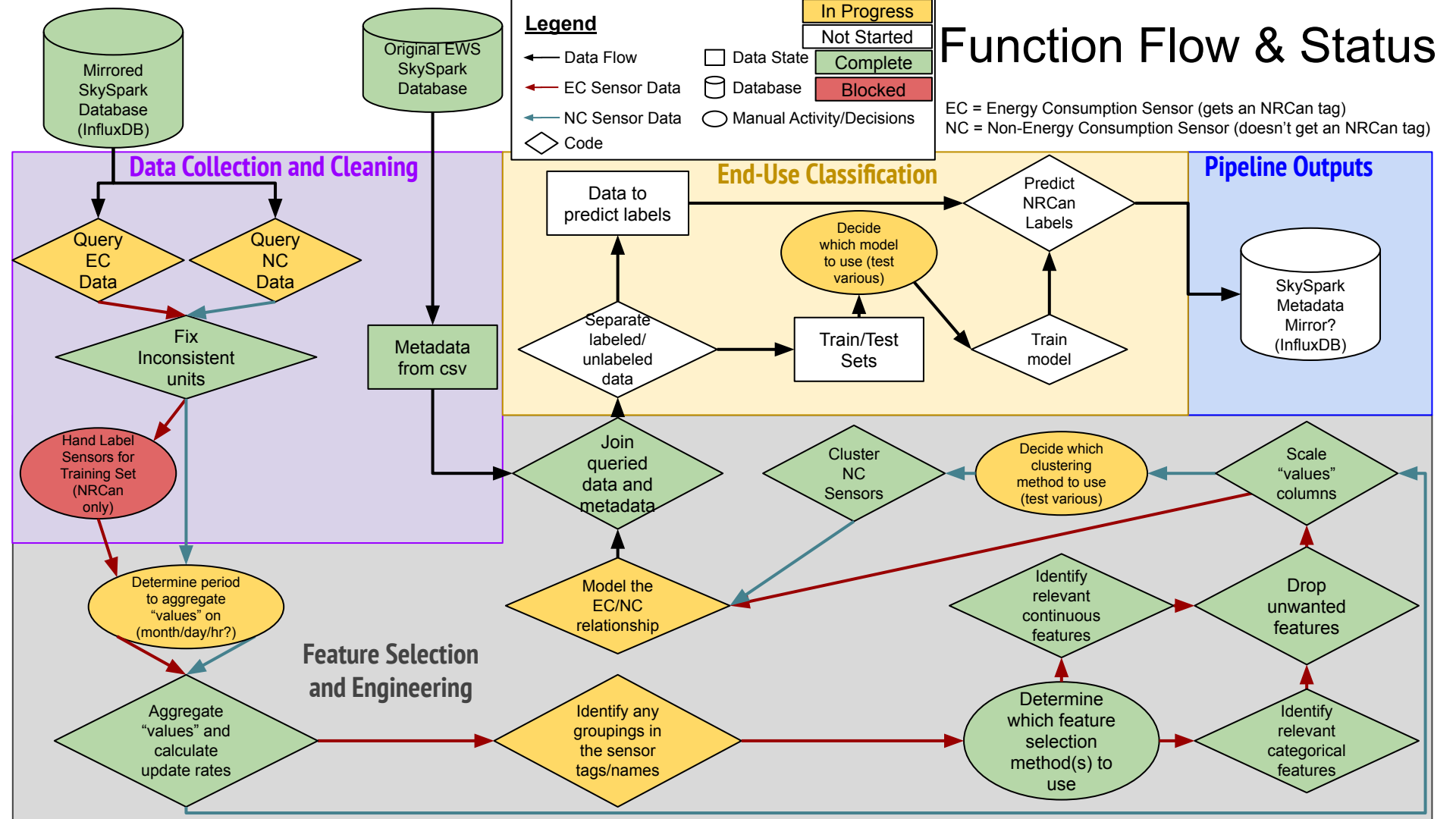
- Adjust parameters of model

Finalize Model & Visualization

- Validate & evaluate model
- Create visualization of results

Wrap-up

- Presentation
- Final report
- Package final code


Function Flow & Status
Legend
Data Flow
EC Sensor Data
NC Sensor Data
Code
Data State
Database
Manual Activity/Decisions
In Progress
Not Started
Complete
Blocked
EC = Energy Consumption Sensor (gets an NRCan tag)
NC = Non-Energy Consumption Sensor (doesn't get an NRCan tag)
Mirrored SkySpark Database (InfluxDB)
Original EWS SkySpark Database
Data Collection and Cleaning
Query EC Data
Query NC Data
Fix Inconsistent units
Metadata from csv
Hand Label Sensors for Training Set (NRCan only)
End-Use Classification
Data to predict labels
Predict NRCan Labels
Decide which model to use (test various)
Separate labeled/ unlabeled data
Train/Test Sets
Train model
Pipeline Outputs
SkySpark Metadata Mirror? (InfluxDB)
Join queried data and metadata
Cluster NC Sensors
Decide which clustering method to use (test various)
Scale "values" columns
Determine period to aggregate "values" on (month/day/hr?)
Model the EC/NC relationship
Identify relevant continuous features
Drop unwanted features
Feature Selection and Engineering
Aggregate "values" and calculate update rates
Identify any groupings in the sensor tags/names
Determine which feature selection method(s) to use
Identify relevant categorical features

Preliminary Results

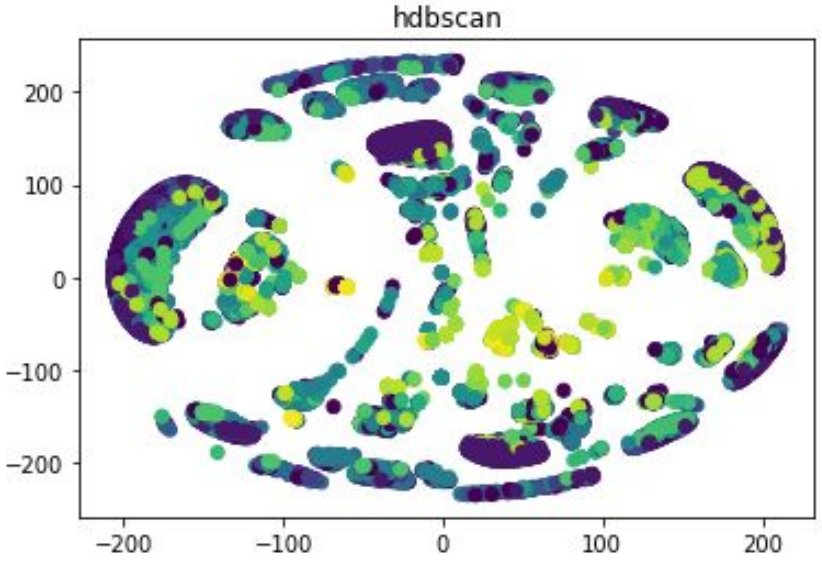
Clustering

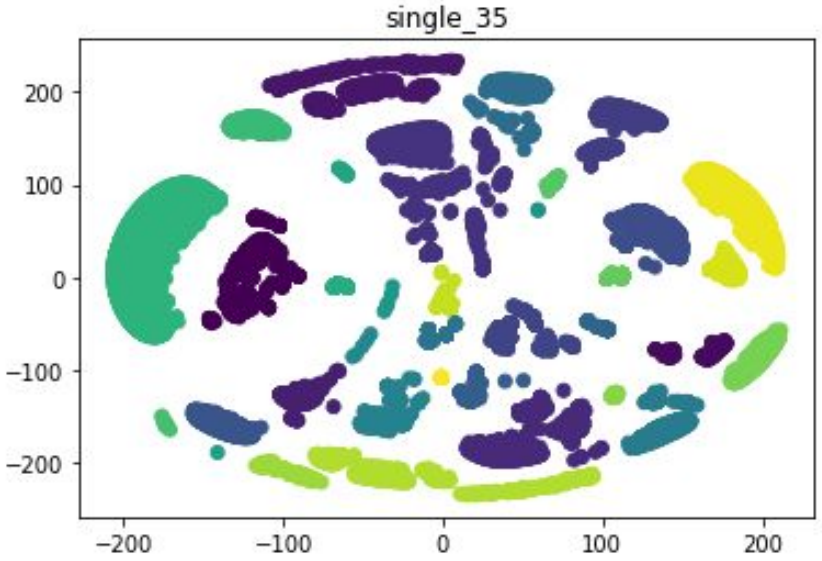
→ Cluster NC Sensors

Date	Hour	mean_0	std_0	min_0	max_0	urate_0	mean_1	std_1	min_1	max_1	urate_1	..._cn
2020-05-01	0	55.2	24.1	0	100	15	10	.1	2	18	1000	
2020-05-01	1	50.1	14.2	5	80	15	10	.1	2	18	1000	
...	...	...	...	...	...	...	...	...	...	...	...	
2020-05-01	23	37	19	1	64	15	5	.1	2	18	1000	

Clustering

→ Cluster NC Sensors





Regression

→ Model EC/NC Relationship

0	1	2	3		17	18	19	uniqueID
0.000037	-0.004377	0.0	-0.000041	...	5.876493	8.502804	20.087383	AHU-01 SF Air Systems Energy AHU1_SF_VFD_PWR(kWh)
0.000039	-0.004622	0.0	-0.000044	...	6.537176	8.851925	20.473544	AHU-02 SF Air Systems Energy AHU2_SF_VFD_PWR(kWh)

Coefficients from Ridge Regression for each sensor

Difficulties & Roadblocks

Difficulties

- Time constraints
- Collaborative coding: making sure inputs of one step match the outputs of a previous step, jupyter notebooks+git, etc.
- Changing and/or optimizing code to work with larger datasets
- Figuring out data range / finalizing dataset
- Effectively explaining details of project to UDL

Roadblocks

- Waiting for response from UDL for diagrams on energy systems
    - UDL wants to be able to identify at least heating, cooling, lighting energy use. We have no labeled example of lighting energy use.

Tasks for Next Cycle

Tasks for the Next Weekly Cycle

1. Develop various classification models for EC Data
2. Add code for each step of the model into the main.py file & make sure they are connected correctly
3. Tune the model → increase accuracy
4. Optimize & clean the code → increase speed

Questions