**Stand-Up Talking Points**

Notes taken by Kaylyn with input from Julia

June 4th, 2021

* Since last week:
  + Added example notebooks, K fold cross validation, and unit tests
  + Started work on developing visualizations in Dash
* What we’re doing next:
  + Working on presentation, dash visualizations, improving/adding unit tests, and adding better documentation
  + Meeting with our sponsor on Monday to go over our final presentation
* Problems:
  + Last set of data has been complicated to generate
  + The visualizations we’ve set up in matplotlib are really good but now we have to figure out how to put them into Dash

May 28th, 2021

* Since last week:
  + Met with the sponsor to get feedback on our ideas for the final deliverables of this project
  + Started work on developing proper visualizations in Dash
* What we’re doing next:
  + Working on final deliverables, improving/adding unit tests, and adding better/reorganizing documentation
  + Filling in analysis that has been requested from sponsor
* Problems:
  + Still not sure about the validity of the model, but it’s what we have
* Dave’s suggestions:
  + Ask Josh and Wes about what they think of our final project [😬](https://emojipedia.org/grimacing-face/#:~:text=Emoji%20Meaning,Apple%20iOS%20prior%20to%202017.)

May 21st, 2021

* Since last week:
  + All models are working on our standard dataset
  + Visualizations have been created to view accuracy of all three models
  + Done initial comparisons of model performance based on differences in input parameter space
  + Created a presentation to give to our sponsor to highlight the work we’ve done so far and where we see this project going
* What we’re doing next:
  + Presenting to our sponsor today for feedback and questions
  + Working on final deliverable, improving unit tests, and adding better documentation
* Problems:
  + Data generation is slow and tedious
  + Some of the trends between input parameter space and model performance were surprising

May 14th, 2021

* Since last week:
  + All three models are working on our standard dataset without predicting the mean
  + Visualizations have been created to view accuracy of all three models
  + List of necessary datasets to probe input space generated
  + Made a pitch video
* What we’re doing next:
  + Comparing datasets on the different models
  + Hopefully being able to present our models and conclusions about interesting input space to our sponsor
* Problems:
  + Data generation is slow and tedious
  + Will our models always work well on any provided dataset?

May 7th, 2021

* Since last week:
  + Standardized the process of generating data sets, focused on generating datasets that we can use with the models to compare isotherm used, types of impurities, etc.
  + Generated data to probe the importance of chromatography resin and also the effective adsorption coefficient
  + Generated data pre-processing pipeline and started unit testing
  + Started accuracy metric pipeline and unit tests
* What we’re doing next:
  + Comparing datasets on the different models
  + Models need some updating (see problems)
* Problems:
  + Very simple model is predicting the same value for yield and purity with the current datasets; is this an error with our datasets or is there something wrong with our model?
  + This has slowed down the accuracy metric pipeline.
  + Ctrl-F doesn’t work for editing .py’s in Jupyter

April 30, 2021

* Since last week:
  + Continuing to generate data (larger datasets) using the mechanistic model
  + Met with our sponsor to clarify parts of the model and determine interesting ranges of parameter space that they want to understand
  + Updated Gantt chart to match what we’ve spoken about with our sponsor
  + Built linear regression model, simple NN, and probabilistic bayesian NN
* What we’re doing next:
  + Continuing to generate better data using the mechanistic model
  + Classifying model performance based on differences in data (isotherm used, size of data set, etc.)
  + Cleaning up our data processing and accuracy metric pipelines so they’re standardized across all three models and we can start writing tests
* Problems:
  + Because our data is generated with multiple points from the same experiment, it might be overestimating accuracy since some of the data sets are correlated--working on changing our data processing pipeline so that all points from one experiment are either in training OR testing but not both

April 23, 2021

* Since last week:
  + Figured out how to generate some data (whether it’s good data we haven’t evaluated yet)
  + Literature search on neural network architectures in literature
    - Determined starting models and architectures to focus on
* What we’re doing next:
  + Generating a skeleton framework for NN architecture so we can focus on feature engineering and optimization (already exploring some NNs with dummy data)
  + Starting optimization
* Problems:
  + Making sure everyone has the same base understanding of neural networks so we can start optimization (had a tutorial with one of our sponsors on building NN)
  + Generating accurate data sets with real-world reasonable input parameters

April 16, 2021

* Since last week:
  + Trying to get the available code to work
    - Diego can do this…mostly
  + Other team members have been learning/reading about sampling strategies for this kind of problem as well as what machine learning might be useful once the data is generated
    - The sponsor gave us some great starting points
* What to do now:
  + Get everyone on the same page with using the code given to us
  + Actually use code to generate data
* Problems:
  + Most team members are having trouble using the code given to us by the sponsor
    - We have a meeting later today to address issues/progress with the sponsor.