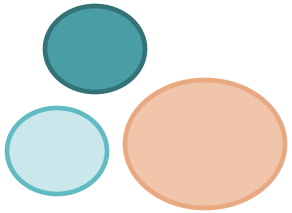
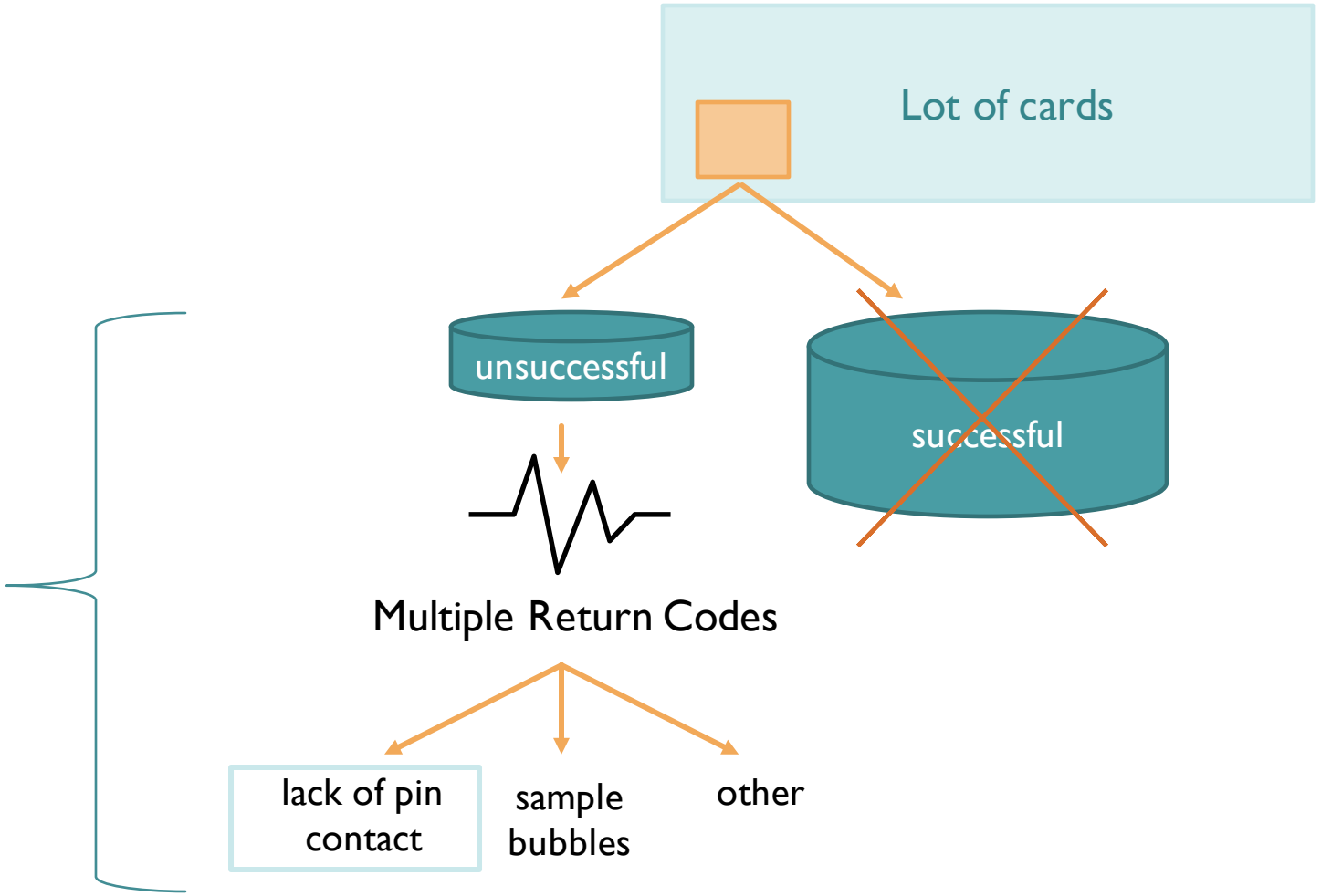
JUNE 14 UPDATE

JUSTINE FILION, NEETHU GOPALAKRISHNA, SAISREE GR, SARA HALL

RECAP: ANOMALY DETECTION IN BIOSENSOR WAVEFORMS

Can we find clusters?





RECAP: ANOMALY DETECTION IN BIOSENSOR WAVEFORMS

Research Questions:

Can we develop machine learning pipelines to cluster readings and isolate pin contact errors?

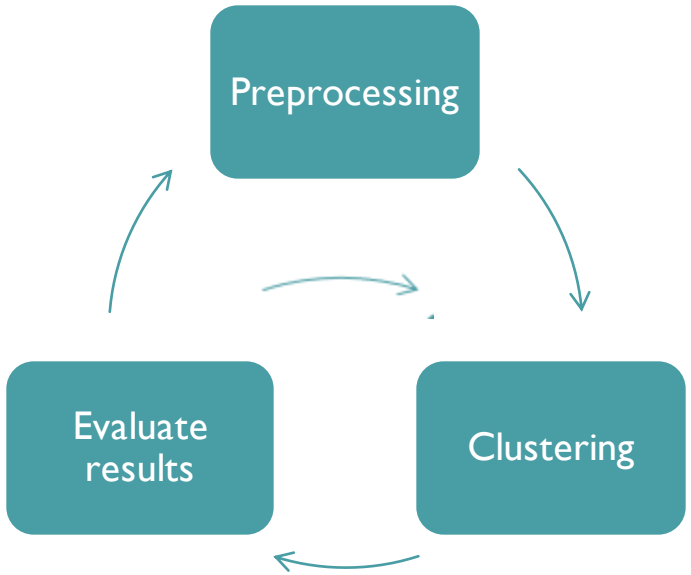
Determine which methods are effective and which are not for identifying anomalies in biosensor readings?

Deliverables:

- Well commented Python code for everything we have tried
  - Preprocessing
  - Clustering
  - Diagnostics
- A final report detailing our attempts
  - Will explain design decisions

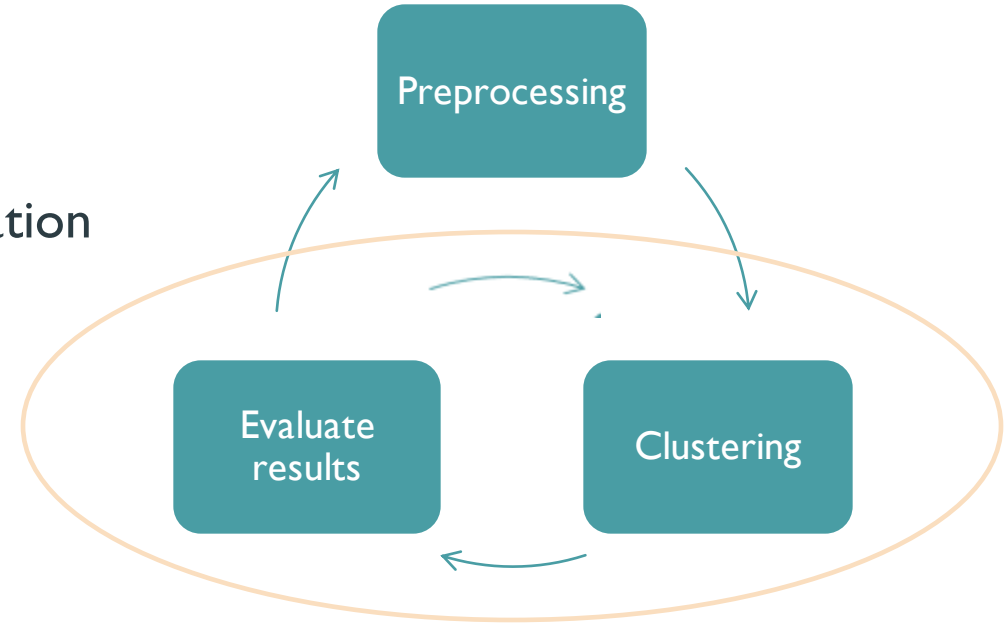
OVERALL PROGRESS

Two Iterations of:



PROGRESS OVER LAST WEEKLY CYCLE

2nd Iteration



Best results

- Autoencoder
- Kshape clustering

RESULTS

CLUSTERING – ITERATION 2

Comparison between 2 promising pipelines:

Autoencoder, PCA (Timeseries), Predictors

- More preprocessing steps
- Clusters well less defined
- 1 cluster with 73 % of all pins

Kshape (Timeseries + Predictors)

- Less preprocessing steps
- Clusters are well defined
- 2 clusters with 78 % of all pins

CODE CLEANING

- Cleaned our notebooks with proper comments and documentation
  - Preprocessing
  - KShape
  - Autoencoder
  - SOM
  - Diagnostics
  - KMeans
  - PCA
  - RandomForest

CODE CLEANING

- Added our iteration 1 and miscellaneous notebook to the glossary and archived them for future reference on client cloud.
  - Pre-processing (Iteration 1)
  - Clustering (Iteration 1)
  - Noise Reduction (Iteration 1.5)
  - Anomaly Detection (Iteration 1)

Archive Glossary

Here is a glossary for the original bits of code used in the first iteration of our project where the results were unsatisfactory.

Preprocessing (Iteration 1)

Preprocessing steps taken in the first iteration. The notebooks are listed in the order that they were used.

Notebook/Script	Description
SeparateLabels	This notebook was used in the first iteration to separate out the unsuccessful and successful predictors into different files.
CheckMissingRecords	Used in the first iteration to make sure there were no duplicate test IDs or test IDs that were present in predictor files but not the time series files or vice versa.
Standardize	This script was used in the first iteration to remove the wet-up period and standardize the waveforms to have mean 0 and stdev 1. The code was janky though so in our final pre-processing notebook an improved function is included for future use.
window	This notebook was used in the first iteration to separate the files into widows relative to sample detect time. It wasn't great though because the indexing was highly dependent on the amount of wet-up removed. A much better function for this (window_after_zeroed()) is included in the final preprocessing notebook.
Filtering	This notebook was used in the first iteration to try various filtering techniques such as low pass and band pass filters on the waveforms.

CODE STRUCTURE

- Updated the file structure to match with the client environment.
- Uploaded all notebooks to client SharePoint for code review.
- Provided the clients with requirements.txt file with information on all the packages used.

CODE REVIEW

- Clients were able to download and execute the code from their end without any issues.

PLANNING AND ACTIONS FOR THE NEXT CYCLE

Sara :

- Work on introduction and preprocessing for the report.
- Conclusion and references section.

Saisree :

- Work on Data, SOM and pipelines for the report
- Cluster evaluation section.

Neethu :

- Work on autoencoder, Hierarchical and Gaussian mixture model clustering for the report.
- Future enchancement section.

Justine :

- Work on background, methods, Kshape for the report.
- Along with PCA and TSfresh section.

Team

- Work on the final report equally.
- Report review circle.
- Make slides for next week's presentations.

ROADBLOCKS

- Capture all our attempts in detail in the final report.

DEVIATION FROM THE ORIGINAL PLAN/SCHEDULE

ACCOMPLISHED ALL LAST WEEK'S TASKS?

- Last week's tasks:

Sara :

- Clean PCA and Preprocessing notebooks
- Waveform characterization
- Make a glossary to organize and describe code

Saisree :

- Clean the SOM notebook
- Work on cleaning up Github repo and making sure it's up to date
- Documenting the work done so far

Neethu :

- Clean the autoencoder and random forest feature importance notebook.
- Try running K-shape with some autoencoder features as additional predictors.
- Create slides for the next weekly cycle

Justine :

- Clean the KShape notebook
- Work on characterizing the clusters using density plots for the predictors
- Upload our preprocessing csv files to the client's cloud
- Change all the directories in the notebooks so that they match the ones in the client's cloud

➡ No deviation we cleaned up and peer reviewed notebooks before uploading it to client cloud.

DEVIATION FROM THE ORIGINAL PLAN/SCHEDULE

AHEAD/BEHIND/ON TRACK?

May 16 - June 5	Modelling	• Try to build various machine learning pipelines to figure out what works and what doesn't in terms of clustering different types of unsuccessful readings • If the unsupervised pipelines are unsuccessful, we will try building some supervised pipelines to classify successful, unsuccessful, and pin contact. • Midterm presentation May 31.
June 6 - June 12	Tuning	• Focus on improving the most promising model(s). • If we have time, maybe look into data augmentation methods.
June 13 - June 19	Documenting	• Write the final report. • Generate slides and practice for the final presentation.

SUMMARY OF INTERACTIONS WITH THE CLIENT

- Exchanged a few emails throughout the week
- Advisory Committee Meeting on Monday, June 6th
  - Gave an update of our progress to the larger team at Siemens Healthineers.
- Meeting on Tuesday, June 7th
  - Gathered information regarding client file structure for setting paths
- Meeting on Friday , June 10th
  - Code review, client downloaded and executed all our notebooks on their machines.

SUMMARY OF INDIVIDUAL AND TEAM EFFORTS MAY 30 – JUNE 3

- Sara :
  - Meetings/Presentations: 3
  - Coding/Debugging/Documenting: 30
  - Administrative work: 1
  - Report Writing: 4
  - **Total : ind. + team = 38**

- Neethu :
  - Feature extraction and modelling :
  - Researching :
  - Administrative work :
  - **Total : ind. + team =**

Saisree :

- Debugging/ Coding/ Documenting:
- Report writing :
- Administrative work:
- **Total : ind. + team :**

- Justine:
  - Coding :
  - Presentations/Meetings :
  - Research :
  - **Total : ind. + team =**

Team:

- **Time spent in meetings : 3 hrs**

NEXT STEPS

1

Write the final report

2

Present our project to the advisory committee

3

Create our final presentation slide deck

QUESTIONS OR FEEDBACK?

THANK YOU FOR YOUR TIME!