Team name

- Agile

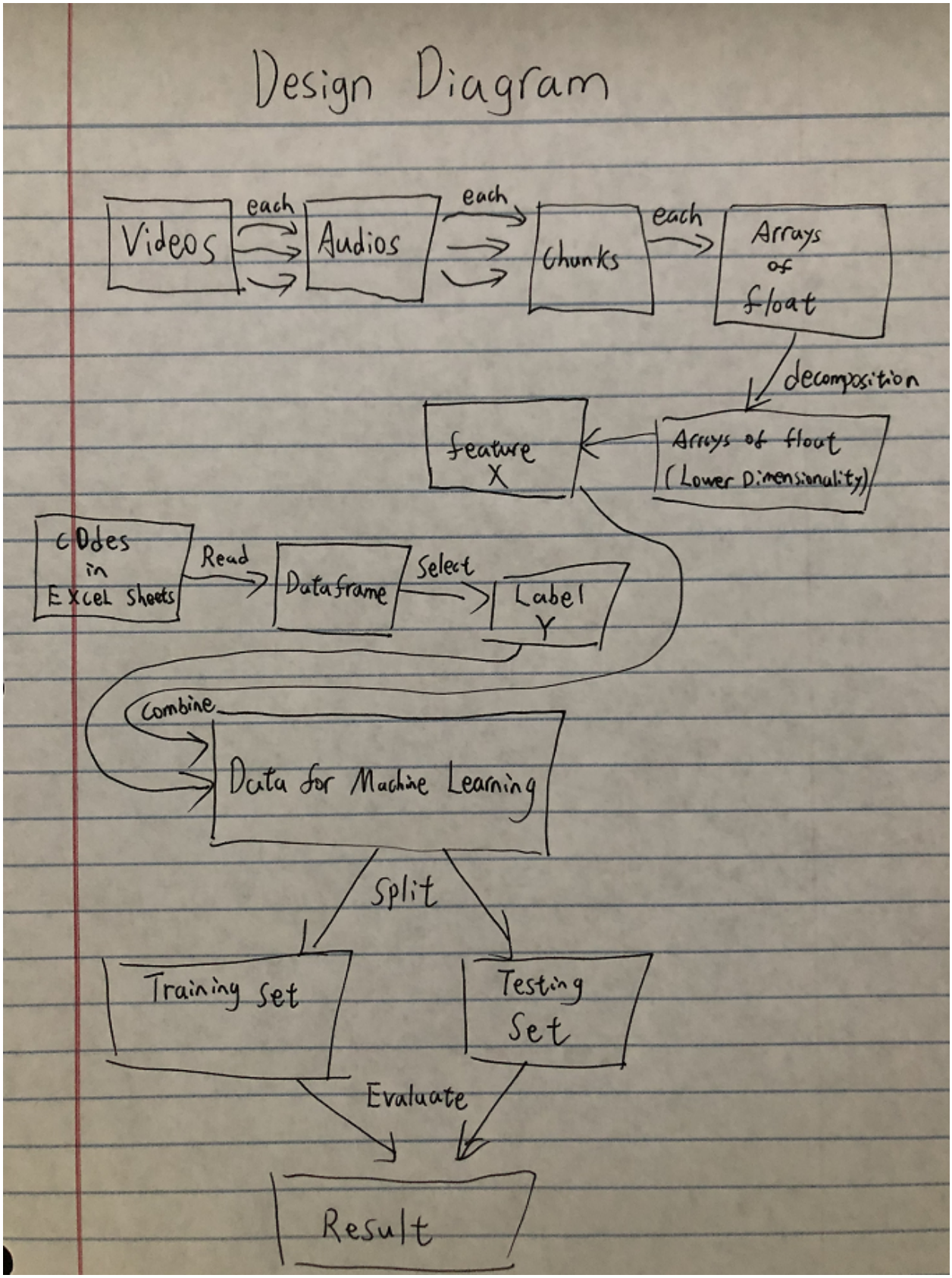
Team roles

- Pei Chen: Developer
- Jin Huang: Developer
- Yuanyuan Lei: Product Owner
- Dongqing Yang: Developer
- Han Zhang: Developer
- Rongruo Zhou: Scrum Master

Customer meeting date/time/place

- First Meeting: 10/23/2021, Saturday, 10:00 am, Zoom
- Weekly regular meeting: every Wednesday (Start on Oct.27), 11 am – 12 pm, Zoom with Cindy, Guerrero, and her group members.
- Weekly regular meeting: every Wednesday (Start on Nov.3), 5 pm - 6 pm, Zoom with Dr. Irby for updates and clarifications
- Second Meeting: 10/21/2021, Wednesday, 11:00 am, Zoom
- Third Meeting: 11/10/2021, Wednesday, 11:00 am, Zoom

Design Diagram


Design Diagram
Videos
each
Audios
each
Chunks
each
Arrays of float
decomposition
Arrays of float (Lower Dimensionality)
feature X
codes in Excel sheets
Read
Dataframe
Select
Label Y
Combine
Data for Machine Learning
Split
Training set
Testing Set
Evaluate
Result

User Stories Finished

- Split the audio from the video. I used a similar approach as in the legacy code. The generated audio file will be stored for future usages.
- Split the audio into 20 seconds chunks. I used a similar approach as in the legacy code. The generated chunks will be stored for future usages.
- For each of the chunks, read the audio file and convert the file into an array of floats. It’s similar to storing the images as arrays.
- For each of the arrays of floats, do the decomposition to lower the dimensionality. Since the previous arrays have an extremely high dimensionality like each array contains about 844,000 floats, which is nearly impossible for machine learning.
- Read the provided excels sheets into pandas data frames. Combine all the data frames into one main data frame.
- Add the edited arrays of floats into the data frame to relate each of the audios to its corresponding codes.
- Using the arrays of floats as X, and the target code as y, for example, I use "Lang of Instruction Teacher" as required by the customer. Split 80% of the data as the training set and 20% of the data as the testing set.
- Using a machine learning module (I currently use SVC) to train and make predictions. The current accuracy is about 68%.

Note

- Many of the user stories have been changed due to the difficulty and negotiation with the customers.

Pivotal Tracker

- https://www.pivotaltracker.com/n/projects/2535924

GitHub

- https://github.com/oniremilia/ILOI

Slack

- https://join.slack.com/t/csce606iloi/shared_invite/zt-xglog0ed-F8eRaqVaKMo0KheFO8DD5A

Custom Grading Request

- Since the ILOI project is a pure machine learning, or natural language processing project. We do not have any requirements from the customers to do the web stuff. The customers are looking for locally running machine learning modules that can generate result forms as desired.
- Hence, I can not deploy to Heroku. But the code can run as described before locally, and each of the user stories can be tested using data similar to the data I’m using. Also, tests like RSpec or Cucumber are not doable.
- The current code completes 50% or above of our customer requirements.