

Weekly Schedule Schedule: Sunday, 2:00 PM

Our github repo: [alckasoc/Team-Chiken-wi22: The official ACM AI Team Chiken repository. \(github.com\)](https://github.com/alckasoc/Team-Chiken-wi22)

3/23/2022 Meeting Time: Wednesday, 2:00 pm

Attendees: Derrick, Min, Vincent

Summary of Meeting

- Discuss 2/27/2022 meeting notes and project clarifications
- Established action items for the coming week

Action Items

By Friday (3/25):

1. Learn easyOCR (figure out how it works) -> google stuff, look at docs, watch tutorials, experiment with the code, etc
2. Go through a basic sentiment analysis pipeline -> you can find this on tensorflow docs (or generally anywhere; this is a pretty popular first NLP project)

By end of spring break (early week 1 of spring qtr):

1. Model your project after the sentiment analysis tutorial
2. Meet up with team Sunday 2:00 PM to collaborate, discuss, and work on project logistics
 - a. Feel free to also work on the team-wide project on your own (but do keep in mind to keep everyone updated on your contributions)!

Timeline

1. Continue (and eventually finalize) technical process
 - a. First point, understand OCR (generally speaking) and understand sentiment analysis with deep NLP
 - b. After you get your fundamentals down, apply your knowledge and learned practical skills to a custom dataset (y'all choose)
 - c. Attach that OCR component into your customized sentiment analysis pipeline
 - d. Maybe tune things, experiment, play around with the code!
2. Expand our horizons and turn this into an app (if time allows)
 - a. Not sure if we have time, but if we do and we have the relevant expertise, then we can definitely beautify this project
3. ACM Project Showcase
 - a. ACM will be having a project showcase (which is ages from now, but still nice to include here)
 - b. Slideshow, visuals, demos, and thorough explanations of your entire project from when you started to the finished product!
 - c. Talk about your difficulties, your many approaches, what you've done, what y'all have discussed, how you fixed issues, what you might do going forward, etc.

Timeline - Technical Process

(This is my sketch of the project)

1. Flush out the problem statement
 - a. Are we using deep learning, if so what aspect of deep learning?
 - b. Is this problem feasible? Will it require rare data?
 - c. Where do we find the data?
 - d. Is this problem suited for AI?
2. Find reliable dataset(s)
 - a. Resources: Kaggle, UCI, other ones available online
 - b. Is this dataset suited for our task?
 - c. What is it missing? Pros and cons?
3. Wrangle the data
 - a. Are there missing values? Are there missing features?
 - b. Are there some underlying problems in the dataset?
4. Preprocess the data
 - a. Transform the data into a format your model will like!
5. Modeling
 - a. Custom model or use off-the-shelf/ready-to-use models
 - b. What model should we be using? Why did you pick that model? Pros and cons?
 - c. Just from learning about how the model works, how do we improve this model?
What are its weaknesses?
6. Inference/Validation
 - a. Inference and validation are different things!
 - b. Validation is tied to the model training loop and inference is getting a few new samples of predictions from the model
 - c. I put both of them here because you monitor your model's performance through validation and also sometimes inference
 - d. Let's see some results and samples.
 - e. Dive deeper into its weaknesses and perform some sort of error analysis!
7. Repeat
 - a. There are many components to your ML pipeline (points 2-6 above) and also many other subcomponents (that may be excluded from the above points)
 - b. Find a concrete way to diagnose the weaknesses of your *pipeline* (not model but the entire pipeline from ingesting the data to making predictions) and adjust it accordingly
 - i. Maybe the model is underfitting or the hyperparameters are wack
 - ii. Maybe the dataset is poor in quality
 - iii. Maybe your preprocessing pipeline needs to be improved
 - iv. etc