JAD1-Tuesday

Team Name: Sushi team

Time: Tuesday 3, 2022 1:00pm

Project: Machine Learning

Description(Scope):

Meeting every Thursday around 1PM with Jared to work together

Project Integration will be determined as

Team Roles:

1. Team leader: Noah
2. Project manager: Juan
3. Documentation: Jordan
4. Other roles:
   1. Programer: Keith
   2. Programer: Miguel

Requirements:

* Python
* Taco
* Git Hub
* Json - Look into/learn reading from + writing to

### Norms of behavior

* On time: Team members will arrive on time. If team members are unable to arrive on time, they will make an effort to notify the team. Absences for emergencies will be excused so long as the team and Professor Wu are made aware of them.
* Respectful: Team members will act respectfully towards each other and bullying, harassment, and similar forms of disrespect are prohibited.
* Willing to learn: Each team member will be asked to learn new information and skills to complete their work on this project. This learning will both be inside and outside of the classroom.
* Open to help: If a teammate is struggling with an aspect of their work, they are expected to ask for help. Similarly when asked for help, team members are expected to assist their teammate so long as they are reasonably able.
* Communication: General team communication will be done via the group’s Discord channel.
* Quality: Team members are expected to finish their tasks at the best quality possible, satisfying the client. If work is considered poor quality by the rest of the team, it will be redone.

### Questions for JAD 1:

* Specific Data set!
* Our understanding of this project is that we will not be interacting with the database, rather we will process a set of images and return information to the UI team who in term will interact with the database. Is this correct?

Noah

* How will we access the server?
* How frequently will we access it?
* Will we be needing an Anaconda Nucleus account for this project? Will the costs be covered by KAB?
* Are there other costs that will come into play?

Jordan

* What basic metrics do we want in the formatting of our output?
* We could break it into the super category/

Miguel

* What/how will the GUI output be formatted for us to work with (what parameters we will be using from the GUI)
* Are we using a database? If yes, what type of database we will interacting with? What information will the database stores?

Keith

* What packages will we need to install with Anaconda and how should we configure our environment.
* How was the existing ML Algorithm handling images, does it deal with “Trash Objects” that are not currently trash. I.e. a cup in someone’s hand.

Juan

* Would the primary audience be solely towards the organization Keep America Beautiful?
* How would we handle image distortion or poor image quality when it can’t be determined what an object is?

Well this is primarily if we don’t go through with TAco:

* When it comes to asking Shultz for having a set dataset showing what is trash and what is not trash so we can easily implement the updated machine learning algorithm what would be the best way to go about communicating this with him?
* Draft questions: bring up the need for a fixed data set so that can be used to implement the machine learning more smoothly.

Juan- Goals for understanding

Which group of pixels represent trash minimum? (target now)

Next task

What type of litter is this? EX: bottle can also how many of them

After come up after

## Notes

* Need packages to work + anaconda environment
* JSON needed for the Input/Output
  + Input is images that will be separated to into different sets
  + Output is text info about the litter detected in the photos

Questions

Goals: (General overview from starting to endpoint)

**THIS PAGE IS STRICTLY to brainstorm questions/matters we’d like to ask Dr. Schultz:**

**JAD#1 Question Brainstorming (Delete after final list is decided)**

**● Ask Dr. Schultz for confirmation of previous groups Github account. Litter Detection**

**○ Previous group members: Chris B., Kathy M., Greg S., Malia T., Andrew Z., Vanessa J. from 2018 CIS 490 class.**

**● Since the project is looking for a fairly high degree of granularity can we agree to start with just the city of San Marcos? If we can get it working for a single city then it shouldn’t be much more work to expand it to the county level. (This would have to be confirmed with KAB UI group so they know that the dataset will only exist for San Marcos, but that they should design a modular UI so that data can be aggregated for each city.)**

**● There are many metrics that could be derived from the ML algorithm and data that is available for each city. As an example: (If anyone can think of other metrics please add them here, also please double check my logic on calculations.)**

**○ City of San Marcos Area: 24.36 sq mi (128,620.80 sq ft)**

**○ CSM 2018 Population Estimate: 96,847**

**○ (Hypothetical) ML Algorithm Pieces of Litter: 21,000**

**○ (Hypothetical) ML Algorithm Images Processed for CSM: 10,000**

**■ (Metric #1) Litter per sq ft: 21,000 / 128,620.80 = .1633 pieces of litter per sq ft.**

**■ (Metric #2) Litter per person: 21,000 / 96,847 = .2168 pieces of litter per person.**

**■ (Metric #3) Litter per image: 21,000 / 10,000 = 2.1 pieces of litter per image.**

**■ (Metric #4) Litter Saturation: Litter per sq ft ( .1633) \* 100 = 16.33%**

**○ Are any of these metrics especially appealing to Dr. Schultz? (Note: Since street view images only account for the area where roads exist it’s not a perfect measurement when using a city’s total area. Also, population and landmass would either have to be hard-coded into a table entry for that city or somehow programmatically pulled from another data source.)**

**● Will we be running this system to be entirely on AWS? (Jerry)**

**○ How shall we go about setting it all up on AWS?**

**○ Any constraints/Issues with using AWS(e.g. Slow server)?**

**○ Technical constraints/pricing constraints (Matthew)**

**● My Current understanding of the general steps we will need to take (Jerry)**

**1. Collect data (Google Street Images) with the use of google API**

**2. Format this data such that it will work well when fed into ML System**

**3. Capture the output (Numbers that categorize the level of litter)**

**4. Push all results to a designed database (that will in turn be used by the UI team)**

**● How frequently do we want the data to be updated ? (Jerry)**

**-maybe pull annually**

**● Give our idea of the project and ask if that represents what he was expecting.**

**●**

**● Now that we have a better understanding of the project, what is your exact vision for this project (Enrique)**

**○ (I know he wants a quantitative number) But from my understanding, the program runs a picture through rcnn, and once the objects are found and marked as litter, there is a quantitative number that is returned 1-4, (4 being more than 9 objects of trash in the picture) but what quantitative number does he want? Because their algorithm already counts how many objects are detected as trash, and just rates the amount of trash in the picture.**

**● What are your must haves, and wants for this project. (Enrique)**

**● We would need a database to store the quantitative number, would the school give us access to theirs? Or how would we go about acquiring one. (Enrique)**

**● How did the previous team get their access to the aws server? The ubuntu server they used previously is not free.**

**Project Manager**

****

**Juan Gonzalez**

**Documentation**

**Jordan Fisher**

****

**Keith Gross**

****