

MyAURA: Personalized Web Service for Epilepsy Managements

By: Shiv Patels

PhD Mentor: Aehong Mins

Research Goals

Clean and annotate data gathered from social media sites. Annotated data will be used to train machine learning pipelines to get better at recognizing social media posts with a medical context.

Challenges

- There were not too many challenges with the task itself. I did find it hard to get working in a group environment due to not having past experience. I do appreciate the team as they made me feel welcomed.
- The one challenge with this task was working with social media posts which were written in a different language. I overcame this challenge by employing online translation tools.

Contribution to overall goals

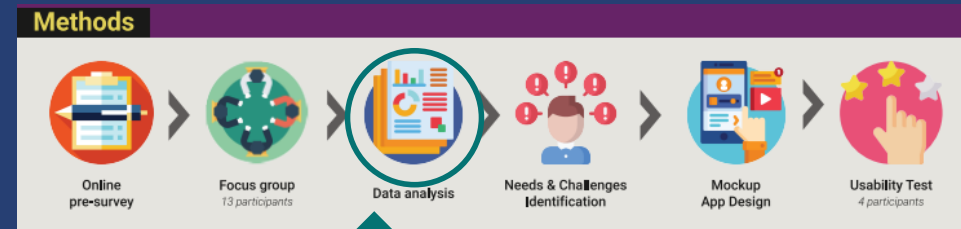


Image by Aehong Min

Current objective

At the project's current phase, my research work will help the researchers train a more accurate ML model which should be able to detect and gather data that is in context, in terms of medical context. While annotating my current datasets, I did notice that with past annotations, the model was able to improve and discern out of context data but it still has ways to go. I am hoping my contribution can help improve it ever so slightly.

This improved model will then be used for creating knowledge graphs for users. Data collected will act as a recommendation system to users.

Methods

The annotations, along with collected data, were stored in Excel files. An example of data is shown below:

```
home made ice coffee , just how i like it .. ☺ .. #home #homemade  
#homesweethome #ice #coffee #justhowilikeit #icecoffee #nespresso #instadrink  
#drink #cold #cappuccino #almondmilk #almond #milk #starbucks  
home made ice coffee , just how i like it .. ☺ .. #home #homemade  
#homesweethome #ice #coffee #justhowilikeit #icecoffee #nespresso #instadrink  
#drink #cold #cappuccino #almondmilk #almond #milk #starbucks
```

Complete image at: <https://imgur.com/a/7cpfgtgs>

Medical term	Nasopharyngitis	0:24-25	0 False-positive Token
Allergen	Almond	0:27-28	1 True-positive Token,

The images need to be split to fit window but each row in both images are supposed to be combined. The method for classification was to read the data collected and determine if the context of the highlighted word matched its classification. Classification is shown as "allergen", "medical term", etc. Putting a "1" would show that the ML model gave the correct classification while a "0" would show that the model did not correctly classify.