

## Functionality

### Run Classifier

“Run classifier” allows the user to directly use our classifier trained on the Reddit dataset to determine the drug addiction recovery propensity of a drug user. The user can choose to use the pre-trained classifier for a single user or for a batch of users. The results for a single user are available on the webpage, whereas for a batch of users the user will have to download the results from the “Results” tab.

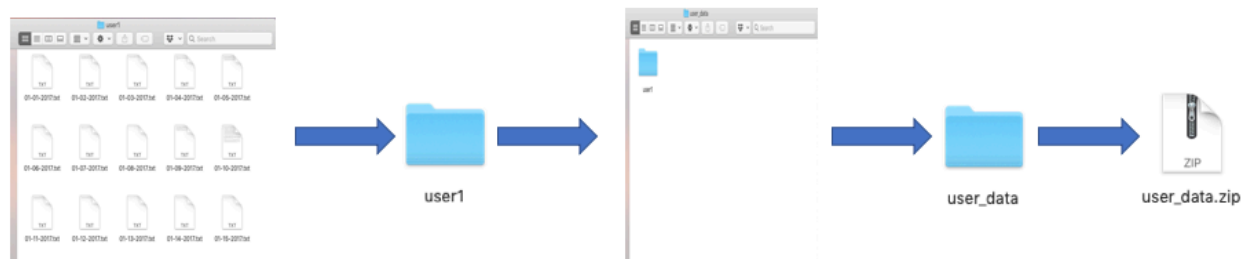
### Train Classifier

“Train Classifier” allows the user to train a classifier on his/her own dataset. After training is complete, the user can download the classifier, and the supporting pickle files from the “Results” section.

## Required Data Upload Format

### Run Classifier – Single User

The upload data for a single user should be a zipped folder. The zipped folder can be named anything, we usually name it “user\_data.zip”. The folder in “user\_data” can have any name, we typically think of the name of the folder as the name of the drug user, “drug\_user\_name”. The “drug\_user\_name” folder should consist of the drug user posts. The name of the posts should be the dates the posts were published and have a “.txt” format. The name of an example post is – “01-01-2017.txt”.



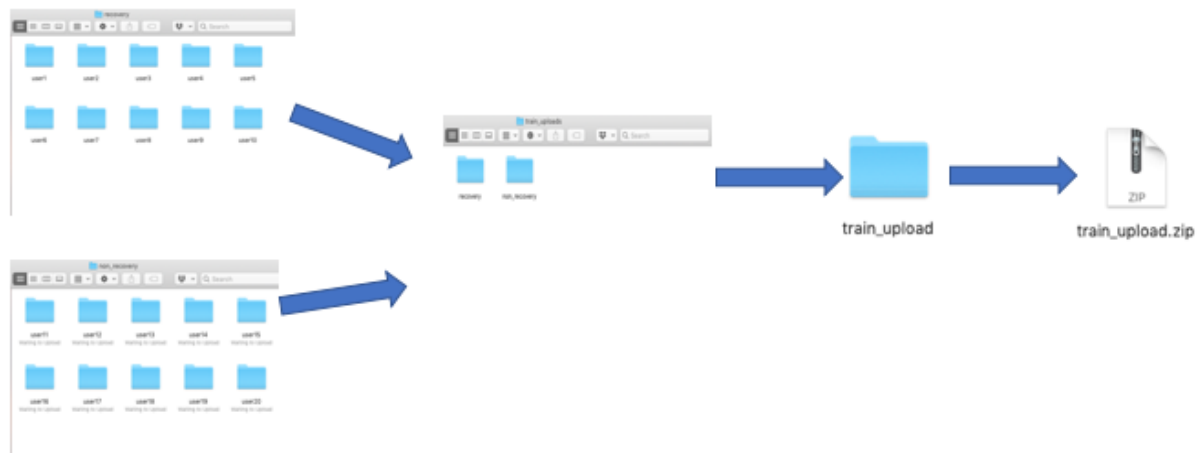
### Run Classifier – Batch

The upload data for batch processing should be a zipped folder. The zipped folder can be named anything, we usually name it “user\_data.zip”. The “user\_data” folder consists of  $n$  ( $n$ = number of drug users) “drug\_user\_name” child folders where each “drug\_user\_name” folder consists of the posts of different users. The “drug\_user\_name” folder should consist of the drug user posts. The name of the posts should be the dates the posts were published and have a “.txt” format. The name of an example post is – “01-01-2017.txt”.



## Train Classifier

The upload data to train a classifier should be a zipped folder. The zipped folder can be named anything, we usually name it “train\_upload.zip”. The “train\_upload” folder consists of two folders “recovery”, and “non\_recovery”. “recovery” folder consists of the posts of the drug users who show signs of recovery. “non\_recovery” folder consists of the posts of the drug users who do not show signs of recovery. Both “recovery” and “non\_recovery” folders can have  $n$  ( $n$ = number of drug users) “drug\_user\_name” child folders where each “drug\_user\_name” folder consists of the posts of different users. The “drug\_user\_name” folder should consist of the drug user posts. The name of the posts should be the dates the posts were published and have “.txt” format. The name of an example post is – “01-01-2017.txt”.



## Example Datasets

- **Run Classifier:**
  - Single user dataset
  - Batch processing dataset

- **Train Classifier:**
  - Train classifier dataset

### **Train Classifier**

Select “Train Classifier” on the home page.

Upload data according to the data format explained above.

Click “Submit”.

Save the project-id.

Enter the project-id in the download results bar to download a zipped folder containing the pickled classifier and supporting pickled files.

User will have to upload the zipped folder when using the trained classifier to predict the recovery propensity.

### **Run Classifier**

#### **Single User (Pre-trained classifier)**

Select “Single User” from the “Run Classifier” drop down menu.

Upload data according to the data format explained above.

Click “Submit”.

Wait for the results to load.

#### **Single User (User trained classifier)**

Select “Single User” from the “Run Classifier” drop down menu.

Upload data according to the data format explained above.

Upload the zipped folder obtained via “Trained Classifier”.

Click “Submit”.

Wait for the results to load.

#### **Batch of users (Pre-trained classifier)**

Select “Batch” from the “Run Classifier” drop down menu.

Upload data according to the data format explained above.

Click “Submit”.

Save the project-id.

Enter the project-id in the download results bar to download the results.

#### **Batch of users (User trained classifier)**

Select “batch” from the “run classifier” drop down menu.

Upload data according to the data format explained above.

Upload the classifier pickled files.

Click submit.

Save the project-id.

Enter the project-id in the download results bar to download the results.