

Problem 1

Approach for social media post segmentation

1. Dataset

The dataset will not be in a single file. Each post will have an id. The texts from the post can be in a csv file while images and videos can be in separate folders having tagged with the post id.

2. Preprocessing

Standard or model wise required preprocessing will be done first for each kind of media in the post. For example, texts will go through lower casing, stop words removal, images will be rescaled etc.

After that the data will be vectorized/converted to numerical form. All different media will be in different feature vectors. At this point there will be weights divided among them. So, a post with text, image and video will have an equally divided total weight of 1. A post with just an image will have a total weight of 1 assigned to that image.

3. Training ML Model

We will be needing summarizer models for texts, caption generator models for images and videos. Each model will be trained separately. Once done our model can generate tags for a post and can even segment it by tags in text, image and videos separately.