

[Link to github code](#)

**1. What are the various options to crack this problem? Any solutions available in the market?**

Ans : To solve the problem text data is available which is very messy and unlabelled as well

There are three types of approaches for solving this problem

- a. Rule based solution : Define the rule for finding action in the text and based on this replicated it for all the sentences
- b. Deep learning based solutions : Since in this case we don't have trained labelled data this cannot be directly used
- c. Hybrid approach : In this method define label for some of the sentences and then based on that predict it on other remaining text docs. This way we can find actions objects for all the text docs

Note: I did not find any date in the text docs so I did not create the label for time. I only did it for actions object.

**2. What tech and approach will you be using?**

Ans : I am using hybrid approach

Steps are like

- a. Cleaning the dataset : since there are too many columns in dataset. I combined all of them and then cleaned the dataset using regex
- b. Annotation of the dataset : since the data is unlabelled. I took out random sample of 120 rows from dataset and then manually annotated the dataset for action labels.
- c. Training: I trained this annotated dataset on the spacy roberta transformer model mentioned in the training\_and\_testing.ipynb
- d. Predicting : The remaining points were predicted on this trained model

**3. What are the challenges?**

Ans : The main challenges faced are

- a. Cleaning the dataset
- b. Labelled dataset is not available : need to manually annotate

**4. What accuracy can be achieved in the first version and timelines?**

Ans: P,r,f observed for test data is 0.9836660617, 0.9944954128 , 0.9890510949

It is not possible to find out the p,r,f on the test dataset because we don't have any labelled data for other points

**5. Solution should have an API which can be invoked from any platform?**

Ans : I did not create an API for right now but to fetch it from any platform the best API I could think of is flask.

**6. How will you secure API?**

Ans : To secure API

1. Create entry barrier and allow from one port only
2. If it is deployed on cloud create a load balancer and allow entry through load balancer only and close connection through any other way.

[Link to trained model](#)

Data\_with\_actions.csv contains the file of actions for the text data