Problem Discussion and assignment of tasks

6th April 2023

Zoom Meeting

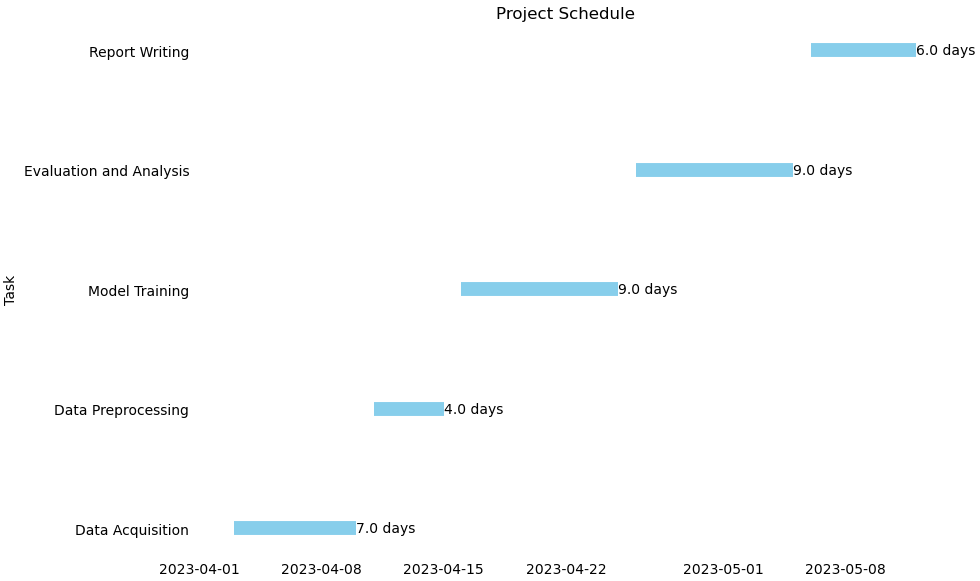
<https://utsmeet.zoom.us/j/89850254582>

**Agenda for the meeting**

* Data Acquisition
* From where would you acquire the data?
* Text Preparation
* What kind of cleaning steps would you perform?
* What text preprocessing step would you apply?
* Is advanced text preprocessing required?
* Feature Engineering
* What kind of features would you create?
* Modelling
* What algorithm would you use to solve the problem at hand?
* What intrinsic evaluation metrics would you use?
* What extrinsic evaluation metrics would you use?
* Deployment
* How would you deploy your solution into the entire product?
* How and what things will you monitor?
* What would be your model update strategy?
* Any preferences for any problem types?

We decided on Data from Kaggle Community on Cyber bullying- <https://www.kaggle.com/datasets/saurabhshahane/cyberbullying-dataset>

* Analyzed what NLP preprocessing to be made.
* Came up with a temporary timeline to finish the task



**Temporary Timeline**

13th April 2023

Zoom Meeting

<https://utsmeet.zoom.us/j/89850254582>

**Agenda of the meeting**

* Working on the challenges while preprocessing the data??- **Faced issues while writing reGex for removing repeated characters and stop words**
* Deciding whether we focus on hate speech/aggression/ offensive language dataset??- **Mix of all as it is cyber bullying**
* Deciding on Models ???

***Decided to work on TF-IDF, Doc2Vector, Counter Vectorisation***

***Tried on random forest, Naïve Bayes, SVM, Logistic Regression***

* How to apply sentiment Analysis??

22nd April 2023

Zoom Meeting

<https://utsmeet.zoom.us/j/89850254582>

**Agenda for the meeting**

Working on the issues faced while implementation of model??

* Addressed all the issues faced by the team members and brain storming session on each algorithms was carried out by each team member to understand the concept and then work on the models.
* It was fun meeting to get to know each other and helping out the other team members so that they can work individually on the models to make sure everyone understand the significance of team work in the group project.
* This week’s issues???
  + Understanding the ways to tackle imbalance dataset
  + What metrics are useful or the project that we are working on
  + How to tune the parameters??
  + How to understand which metrics to use???
  + **Deadline:** when should we have the final report ready??

2nd May 2023

Zoom Meeting

<https://utsmeet.zoom.us/j/89850254582>

**Agenda for the meeting**

* Tracking progress of working on the models by each team members
* Addressing any issues if any by team members
* Evaluation of Metrics
* Assessing all on the models to find the final model

May 8th 2023

Zoom Meeting

<https://utsmeet.zoom.us/j/89850254582>

**Agenda for the meeting**

* Collating all the team members work in one single notebook
* Research and gather data to Start to work on the final report
* Working on writing the individual contribution – ***In our group project, we worked on round robin basis so everyone gets to play all the roles , so it helps him to work on the project end-to-end and understand the significance of each role.***
* Limitations of the project
* Next steps that can be done to improvise the project- ***We as a group could work on few models in our capacity and trying to learn and explore more to work on creating more efficient model for the problem in hand.***

May 13th 2023

Zoom Meeting

<https://utsmeet.zoom.us/j/89850254582>

**Agenda for the meeting**

* Improvising the final report
* Committing the code on github- <https://github.com/prinston27/NLP_AT2>
* We have created a separate branch for test and production – All the initial version are made in test branch – We didn’t track all the progress on github as each were working on small units and we collated and finally committed as single notebook

