



Ahmedabad
University

CSE 523 Machine Learning

Progress Report:

Toxic Comments Classification

Group Details:

NaN-Prediction Pending

Sr. No.	Name	Enrollment Number
1	Aneri Dalwadi	AU1940153
2	Kairavi Shah	AU1940177
3	Nandini Bhatt	AU1940283
4	Mananshi Vyas	AU1940289

Tasks performed in the week:

- Researching about possible approaches to solve our problem
- Visualizing text
- Exploring the data
- Defining classes and plotting graph to see the data for each class available

Outcomes of the tasks performed:

- To start off, the first thing we all needed to do was to learn Apache Spark and it was a great learning experience for all of us and it helped a lot in our project till now.
- Learned about various classification models

Tasks to be performed in the upcoming week:

- Data Cleaning
- Testing out some classification models
- Deciding which layers to use for our model (tf-idf transformer)

References:

1. <https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data>
2. <https://smlltar.com/>
3. <https://www.tidytextmining.com/tidytext.html>
4. <https://arxiv.org/ftp/arxiv/papers/1903/1903.06765.pdf>