



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:

- We made the following requested edits in ppt.
 - Added conclusion about model selection
 - Graphs with specified labels
 - Classification of models used
 - Reasons for selecting 2 character shingles
- We made the following requested edits in the report.
 - Details about the approaches
 - Reasons for implementing various models and selecting non-linear models
 - Added clearly labeled graphs
- We also selected randomized train data and saw how the accuracy varies due to variation in the distribution of input data.
- Chose different dimensions within feature selection using eigenvalues and observed the model performances.

Outcomes of the tasks performed:

- We were able to get clear idea of further approach
- Understand how we could change input data dimensions to observe accuracy of model

Tasks to be performed in the upcoming week:

- Hyper parameter fine tuning of all the linear and non- linear classifiers, including varying and cross validating different numbers of eigenvalues.
- Change the dimension of input data and use of different matrices for dimension reduction.

References:

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

