Hate Speech Detection

Aadithya S (2019115001)
Rohit P (2019115080)

PROJECT ABSTRACT

Project Description:

Hate speech is any form of expression through which speakers intend to vilify or incite hatred against a group or a class of persons on the basis of race, religion, skin colour, disability etc. With social media usage increasing everyday, these networking platforms often become a private space for expressing hate publicly. Our aim is to develop an end-to-end ML project that helps detect and report instances of hate speech on Twitter effectively.

Modules to be achieved:

Complete use of the machine learning pipeline. The processes involved are as follows:-

- Extracting and cleaning tweets using Tweepy
- Exploratory Data Analysis
- Feature Engineering
- Feature Selection
- Choosing and building the best-fit model
- Model Evaluation
- Deployment of the model onto a web application

Potential technologies to be used:

- ➢ *Language:* Python
- ➢ Data Science Libraries - Numpy, Pandas, Seaborn, Sci-kit Learn
- ➢ Various Machine Learning algorithms - used and compared to present the most accurate predictions.
- ➢ Jupyter Notebooks, Google Colab
- ➢ Mercury - web framework useful in converting a python notebook into an interactive web application.
- ➢ Cloud platform to serve the web application (Eg. Heroku)