Toxic Comments Classification

https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge

Project Domain: News / Social Media

Motivation

With the evolution of technology, platforms such as social media that allows the communication of personal thoughts and feelings are increasingly prevalent. However, this degree of freedom is associated with problems such as the promoting hate, hurling abuse anonymously or cyber-bullying - resulting in a toxic online community. Hence, this project aims to study this problem and develop a model that can help identify and classify the various forms of possibly inappropriate comments that stifle productive online discussions.

Statement of the Problem/Task

Develop a model which predicts the probability of different types of toxicity for comments.

Possible Approaches: Traditional Supervised Learning techniques, Linear Regression, Logistic Regression, Decision Tree, K Nearest Neighbours

Key Question (Main goal):
- Can we/our model identify distinguishing traits about comments for use of classification (i.e. use of punctuation, key words, capitalisation)

Additional Questions (Stretch goals):
- Are there patterns between similar comments? (i.e. Are the comments left by the same group of users? How about cases of spam?)
- Find out correlation between outcomes? (Clear that outcomes are not independent since if the comment is severely toxic, it is likely that it would be classified as toxic as well)

General Approach

1. Formulate the key problem to solve and explore the problem scope and
2. Denote problem space
3. Explore possible data analysis tools
4. Engage in basic exploratory data analysis and visualisations
5. Data preparation and clean-up
6. Engineering of features
7. Explore possible classification and predictive models
8. Implement chosen model
9. Review and revise model if necessary
10. Work on stretch goals

Evaluation

Satisfactory Outcome:
- Group works through various relevant kernels / tutorials
  - Kernel for tackling challenge using tensorflow and keras
  - Kernel for tackling challenge using logistic regression (scikit-learn)
- Completes project challenge by week 10 with working solution.

Excellent Outcome:
- Group works through various relevant kernels / tutorials
  - Kernel for tackling challenge using tensorflow and keras
  - Kernel for tackling challenge using logistic regression (scikit-learn)
  - Explore deep learning approaches
- Completes project challenge by week 10 with innovative and improved solution
- Multiple experiments with insights on why final approach was chosen.

Resources

Datasets are provided by Kaggle

https://www.kaggle.com/c/jigsaw-toxic-comment-classification-challenge/data

Schedule / Role Assignment

Week	Tasks
Week 5 (10 - 16 Sep)	• Define and flesh out scope • Complete Project Proposal
Week 6 (17 - 23 Sep)	• Explore and analyse dataset (manipulation and cleaning of data) • Finalise approach to take • Ideation for implementation of the model • Complete peer review of project proposals
Recess Week (24 - 30 Sep)	• First round of implementation • Testing, debugging, improving of first model
Week 7 (1 - 7 Oct) [Midterms]	• Exploration of additional questions and other suitable approaches • Evaluation of current available model
Week 8 (8- 14 Oct)	• Improve on existing model based on what was found out in previous week • Interim Report
Week 9 (15-21 Oct)	• Video (2 Pax), Poster (1 Pax), Report (3 Pax)
Week 10 (22 - 28 Oct)	Lag week
Week 11 (29 - 4 Nov)	• Continue working on Video, Poster and Report.
Week 12 (5 - 11 Nov)	• Continue working on Report.
Week 13 (12 - 18 Nov)	• Prepare for presentation. • Finalise Report.

References

https://www.kaggle.com/sbongo/for-beginners-tackling-toxic-using-keras