Capstone Presentation

By Eric Sundstrom

Natural Language Processing (NLP)

- Programming Computers to Process / Analyze Human Communication

- Popular Applications:
  - Text-to-Speech
    - Helps Visually Impaired
  - Sentiment Analysis
    - Useful in Marketing

Toxic Comment Classification Challenge

kaggle

- Dataset:
  - Comments from Wikipedia Edit's Talk Pages (2004-2015)
- Multiple Labels of Toxicity:
  - toxic
  - severe toxic
  - obscene
  - threat
  - insult
  - identity hate
- Goal:
  - Use Results for Detection/Removal
- Misc:
  - 4,550 teams, $35,000 prize

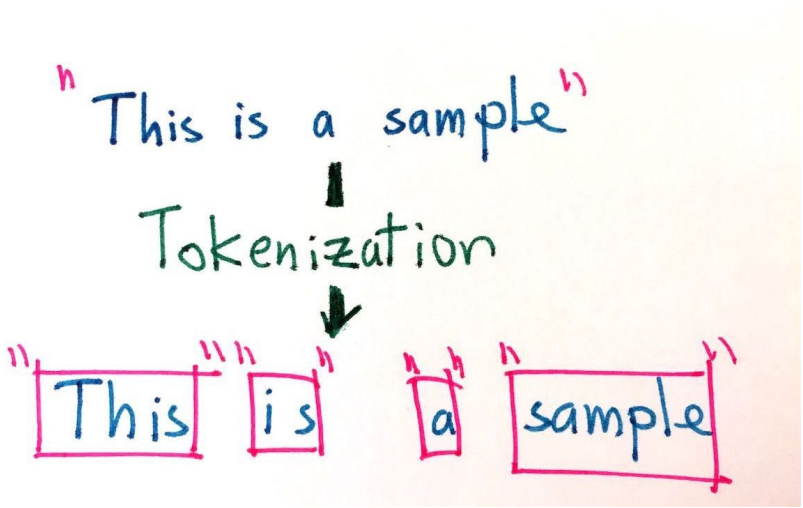
ML Process for NLP

1. Preprocessing
2. Tokenization/Stemming
3. Vectorization (TF-IDF)
4. Modeling

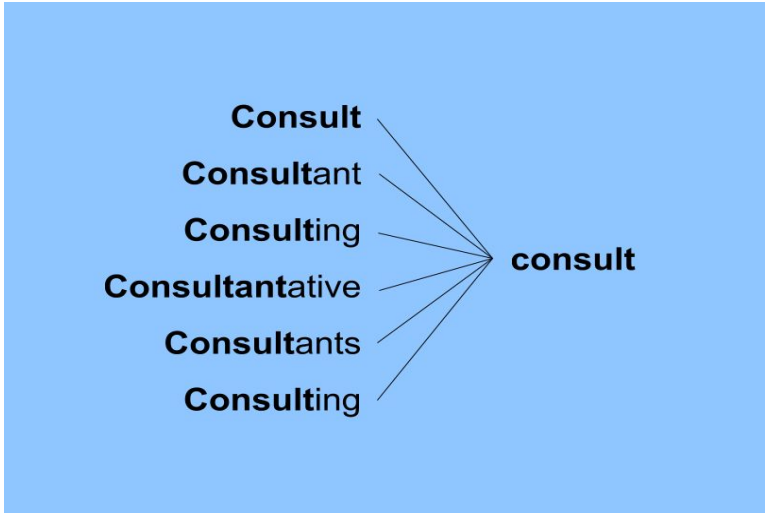
Preprocessing

- Filter Out Clutter
  - Newline characters
  - IP addresses
  - Website urls
  - Domain Specific Abbreviations
    - WP:: __
    - User: __
  - Auto-Generated Text
    - “Preceding unsigned comment added”
    - “UTC”

Tokenization



Stemming



TF-IDF (Term Frequency - Inverse Document Frequency

Vectorization

Toxic Word Cloud

Non-Toxic Word Cloud

Models

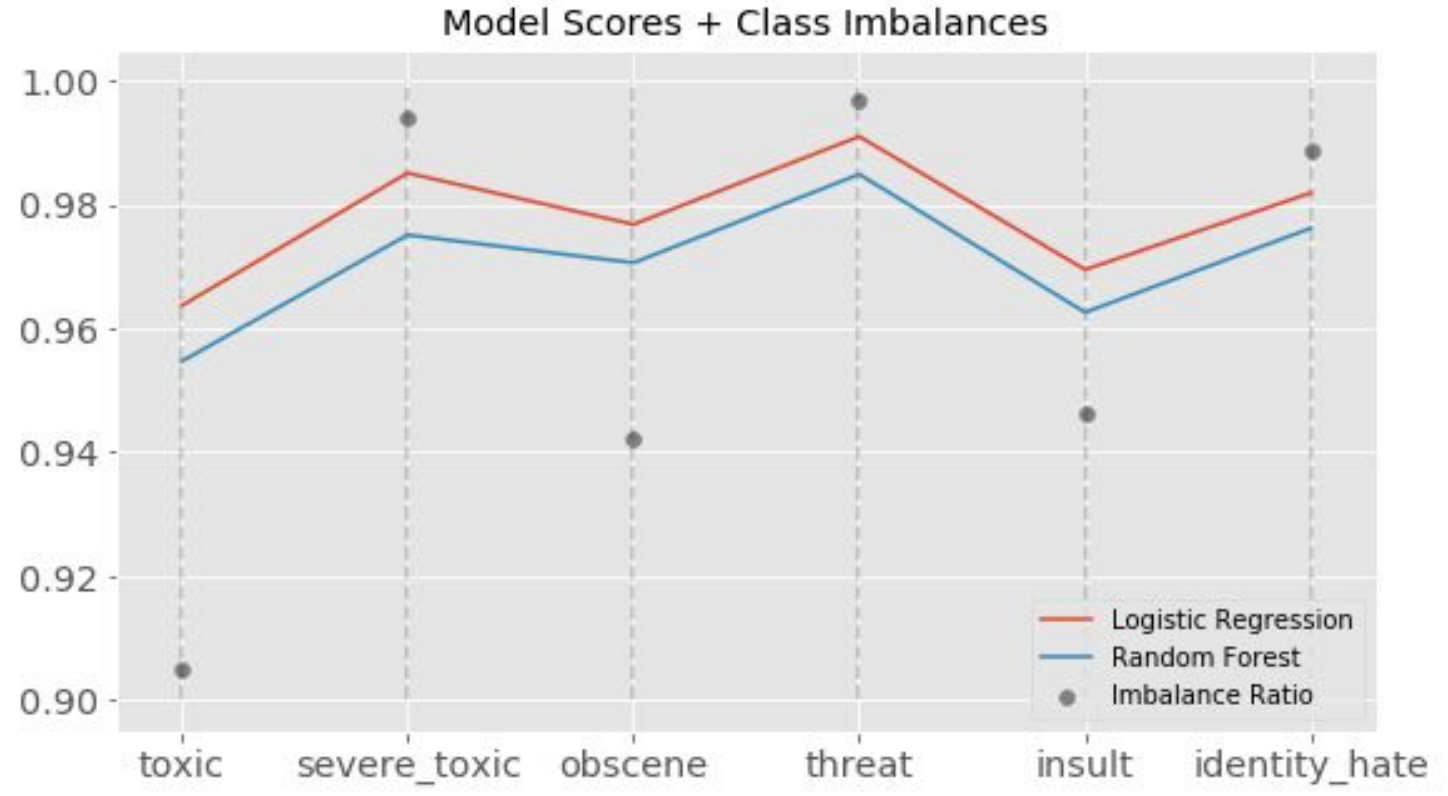
Logistic Regression

- Natural Fit
- Scores
  - Toxic - 96.37%
  - Severe Toxic - 98.51%
  - Obscene - 97.68%
  - Threat - 99.10%
  - Insult - 96.95%
  - Identity Hate - 98.19%
  - **Overall - 97.80%**

Random Forest

- Tree-based Example
- Scores
  - Toxic - 95.47%
  - Severe Toxic - 97.51%
  - Obscene - 97.06%
  - Threat - 98.49%
  - Insult - 96.26%
  - Identity Hate - 97.62%
  - **Overall - 97.07%**

Model Scores Visualization



Results

- Winning Score - 98.86% (171 entries)
- Popularity of Deep Learning
- My Model - 2745/4550 on Leaderboard

Moving Forward

- More Deep Learning
- Use GPU w/ Google Colabs
- New Dataset?

Thank you