Springboard - Data Science Career Track

Capstone Project 1

Identifying Toxic Commentary

By Ellen Savoye - July 7, 2020

Table Of Contents

Introduction

Openly discussing things that you feel strongly about, or care about can be difficult; even more so online or in discussion boards where the threat of abuse and harassment can be prominent. With the advancement of computing technology, including AI, assessment and learning, can a code be developed to recognize toxic comments in online conversations with respect to mentions of identities?

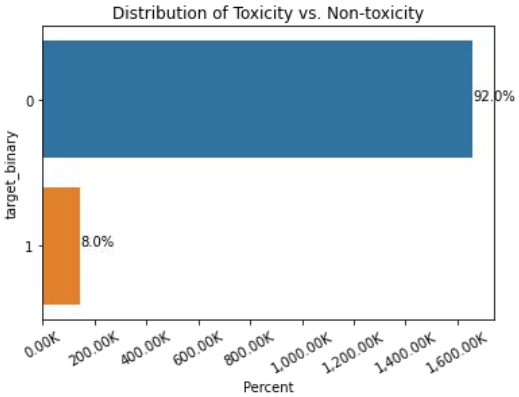
Data

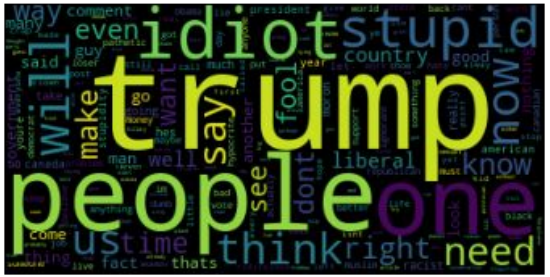
1. Kaggle competition sourced from Civil Comments
2. Train data characteristics
3. Cleaning the data
4. Normalize the corpus using natural language processing

Exploratory & Statistical Analysis

1. Is the data balanced?

2. Toxicity distribution

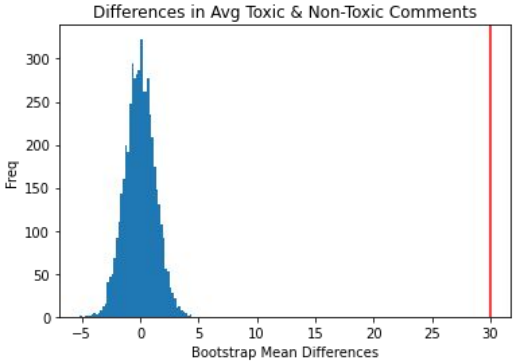


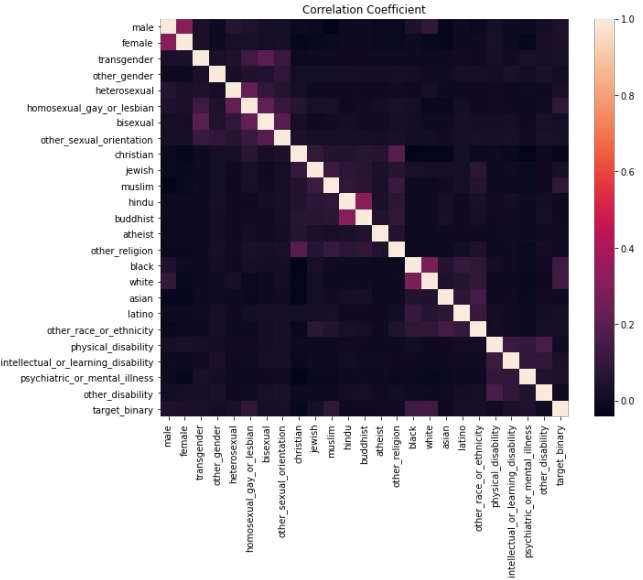


Exploratory & Statistical Analysis

1. Average length of comments

2. Correlation






Predictive
Classifiers

Baseline Classifiers

Logistic Regression

[Training Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.96	0.99	0.97	1245405
Toxic	0.76	0.48	0.59	108250
accuracy			0.95	1353655
macro avg	0.86	0.74	0.78	1353655
weighted avg	0.94	0.95	0.94	1353655

[Test Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.96	0.99	0.97	415135
Toxic	0.74	0.47	0.58	36084
accuracy			0.94	451219
macro avg	0.85	0.73	0.77	451219
weighted avg	0.94	0.94	0.94	451219

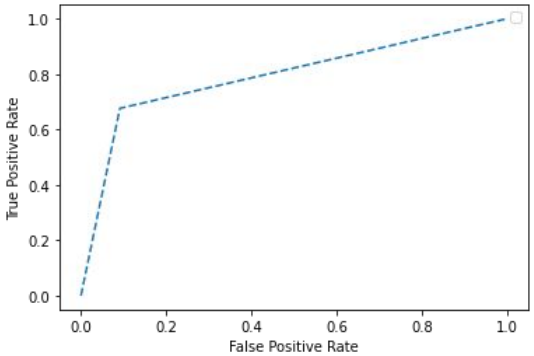
Naive-Bayes

[Training Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.97	0.91	0.94	1162221
Toxic	0.39	0.68	0.50	101190
accuracy			0.89	1263411
macro avg	0.68	0.79	0.72	1263411
weighted avg	0.92	0.89	0.90	1263411

[Test Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.97	0.91	0.94	498319
Toxic	0.37	0.65	0.47	43144
accuracy			0.89	541463
macro avg	0.67	0.78	0.70	541463
weighted avg	0.92	0.89	0.90	541463



Further Analysis - SMOTE

Logistic Regression + SMOTE

[Training Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.88	0.95	0.91	498162
Toxic	0.87	0.74	0.80	249081
accuracy			0.88	747243
macro avg	0.88	0.85	0.86	747243
weighted avg	0.88	0.88	0.88	747243

[Test Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.97	0.94	0.96	415135
Toxic	0.51	0.69	0.58	36084
accuracy			0.92	451219
macro avg	0.74	0.81	0.77	451219
weighted avg	0.93	0.92	0.93	451219

Naive-Bayes + SMOTE

[Training Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.88	0.83	0.85	498162
Toxic	0.69	0.78	0.73	249081
accuracy			0.81	747243
macro avg	0.79	0.80	0.79	747243
weighted avg	0.82	0.81	0.81	747243

[Test Classification Report]

	precision	recall	f1-score	support
Non-Toxic	0.98	0.82	0.89	415135
Toxic	0.27	0.76	0.40	36084
accuracy			0.82	451219
macro avg	0.62	0.79	0.65	451219
weighted avg	0.92	0.82	0.85	451219

Recommendations for the Client

Future Work

1. Regularization
2. `TfidfVectorizer`
3. Undersampling with Oversampling (SMOTE)

Thank you.