A Computational Linguistic Study of Toxicity on Social Media

By Nicholas Turk, Nithya Jayakumar, Toluwanimi Delano, Udai Mallepoola

Toxicity: Common Social Media L

Social media is a collection of human anguish and hate blended together into a horrifying cesspool of our worst impulses as a species.

Detecting Toxicity: A Fool's Errand

Losers on the so-called "cutting edge" of AI technology keep trying to detect toxicity with high-performance classifiers like hamsters running in a wheel of their own creation, unable to comprehend the futility of attempting to diagnose complex social processes.

- **Solution: Reject deep learning, return to linguistics**

Instead of detecting toxicity, we tried to determine the characteristics of toxic language in online writing.

Datasets

Dataset 1: SVM model

- ~3,500 messages from Reddit, Twitter, and YouTube
- Each message classified as toxic or non-toxic by 5 human raters

Dataset 2: NBOW model

- ~160,000 messages from Wikipedia
- Each message classified as toxic, hateful, threatening, or insulting by human raters

SVM

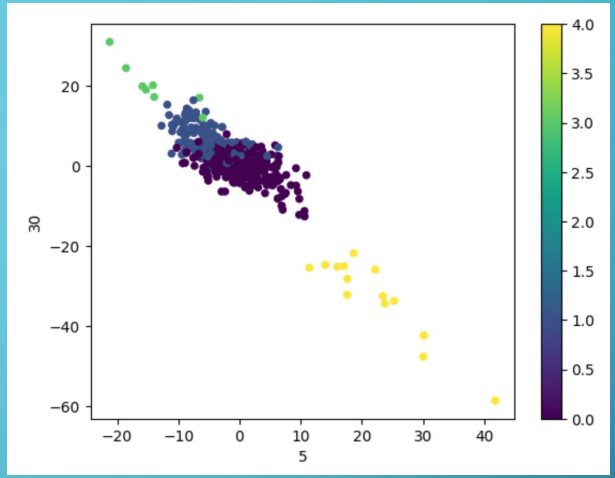
1. Tokenize/process text
2. Extract linguistic features
3. Normalize features
4. Create SVM classifier
5. Determine statistically significant features

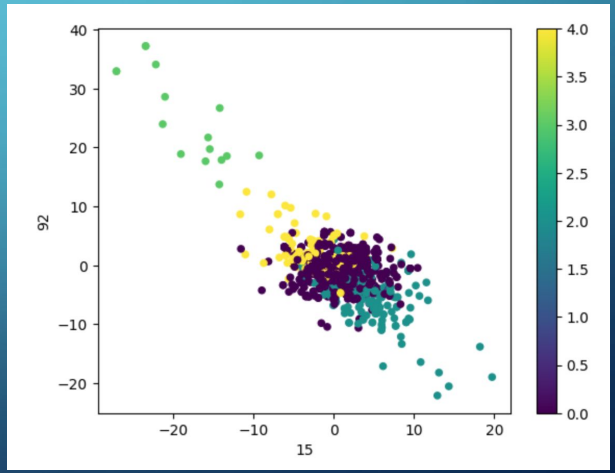
Linguistic Features:

- ❑ Message Length
- ❑ Average Word Length
- ❑ Number of Sentences
- ❑ Number of Misspellings
- ❑ Amount of Punctuation
- ❑ Part-of-Speech Ratios
- ❑ Content-to-Function Word Ratio
- ❑ Type-Token Ratio
- ❑ Length of Noun and Verb Phrases
- ❑ Pronoun Types

NBOW

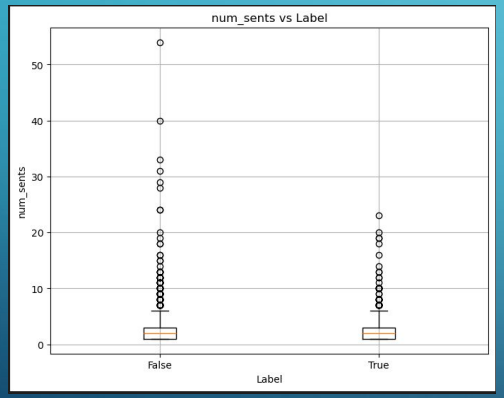
1. Clean and process data
2. Pass data into model
3. Determine similarity and relevance of words by clustering
4. Investigate social and semantic relevance of those words





Linguistic Results

SVM



NBOW

- ★ Second-person pronouns and first-person pronouns appear in different clusters
- ★ Politeness markers clustered together
- ★ Action descriptors clustered together

Questions?

Come at us, bro. I dare you. I double dare you. I triple dare you. I whatever-comes-after-triple dare you. I bet you won't. I bet you're just gonna sit there, complacent, empty-headed. Fine. But don't come crying to me when your presentation comes and no one's there to ask you questions. I bet you're so apathetic, so dissociated and numb that you won't even ask me why this slide's sideways. Maybe you didn't even notice, that's how out of it you are. But you know what, as a reward for having read this far, I'll tell you. The reason this slide is sideways is