
“Wait, was that a joke?”

Identifying Absurdist Humor and Satire

—— Daniel Rossetti Data Scientist
for College University English Dept. ——

Introduction and Problem Statement

College University English Department - studying chat bots

- University Research Goal: *Make them more human*
- Subtask: Help bots identify **humor**

Caveat: Context-based humor:

- **NO** 'setup'
- **NOT** formatted as a joke \Rightarrow Sarcasm, **Satire**, **Absurdist Humor**

Can an NLP classification model be trained to recognize satire or absurdist humor?

Project Outline

- Compare 'News' from subreddits
 - Subreddit: **r/worldnews[1]**
 - Subreddit: **r/TheOnion[2]**
 - Comparable format: Article Titles
- Extract Article Titles
- Create NLP Classification Models
 - Train, Test, Holdout Datasets

Humor Should Be Difficult to Detect

Compare Statements:

- “Supreme Court Overturns **Car**” - The Onion[3]
- “Supreme Court Overturns **Lower Court Ruling**” - No Source, Plausible

Requirements to understanding:

- Word / phrase meanings
- Context expectations
- World knowledge

Sourcing Data and Cleaning

Goal: Isolate Article Title Content

Emojis



UNAUTHORIZED ADVERTISEMENT FOR
COUNTRY CROCK — #nottechnicallysponsored
#SpreadTheCrock 🍷

Hashtags

French President Macron lambasted for throwing
Taiwan under bus | Taiwan News | 2023-04-10
13:06:00

**Useless
Info**

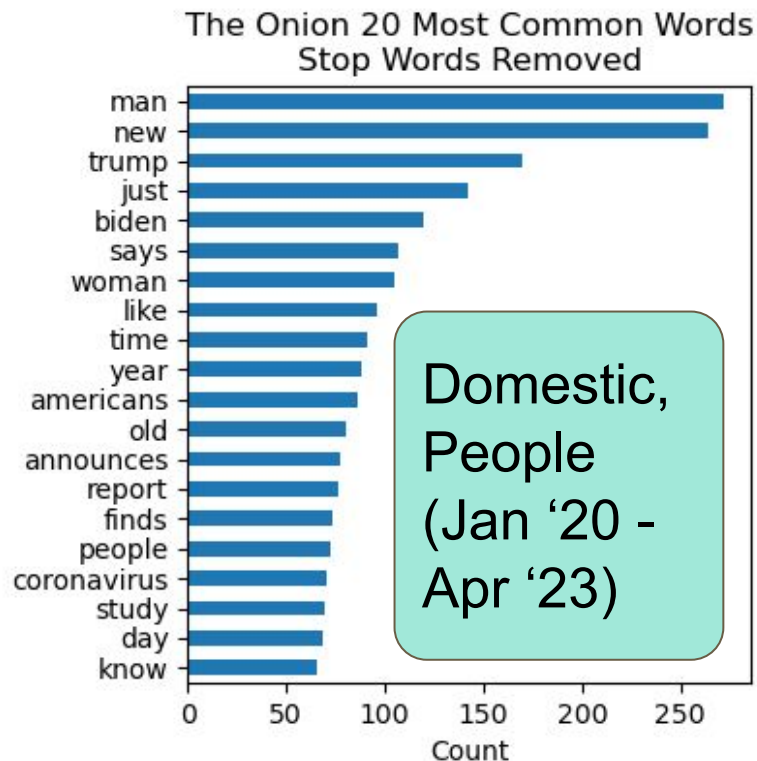
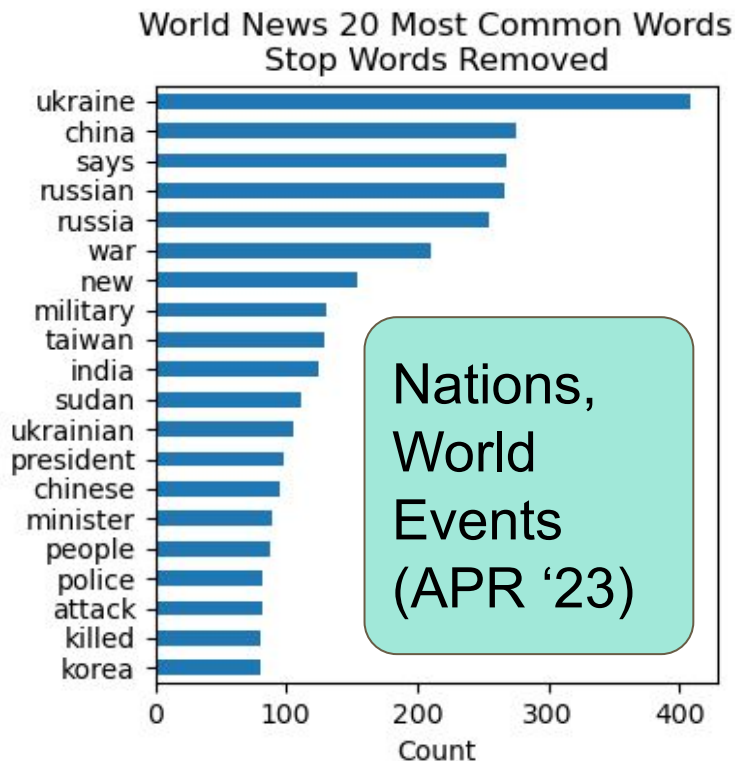
News Agency Names

Model Performances - Accuracies

	<u>Training</u>	<u>Testing</u>	<u>Holdout</u>
Optimized Logistic Regression	99.98%	89.10%	85.99%
Optimized Random Forest Classifier	100.00%	86.67%	82.63%
Optimized Extra Trees Classifier	100.00%	87.91%	84.31%
Stacking Classifier	99.99%	89.35%	86.05%

Accuracy is key: False Negs. / False Pos. Unimportant

Discussion - Subject Matter Makes a Difference



Conclusions

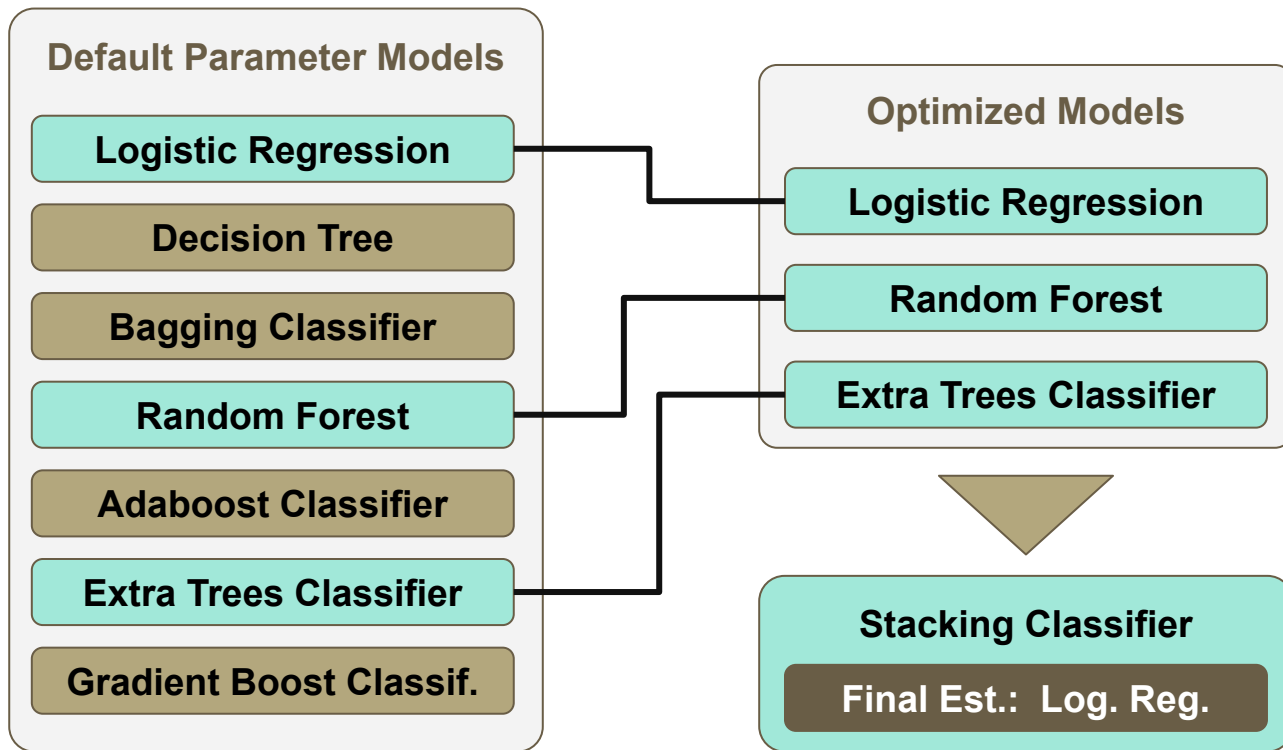
- Model can discern between subreddits
 - Likely due to subject matter
- Model likely unable to understand satire or absurdist humor
- First step towards getting there

Next Steps

- Include subjectivity analysis
- Find other sources of satire
 - With similar subject-matter
- Same time period data
- Try more advanced models

Additional Materials

Modeling Flow Chart



Confusion Matrices

Logistic Regression

		Predicted Label	
		World News	The Onion
True Label	World News	869	128
	The Onion	91	922

Random Forest

		Predicted Label	
		World News	The Onion
True Label	World News	826	171
	The Onion	97	916

Extra Trees

		Predicted Label	
		World News	The Onion
True Label	World News	842	155
	The Onion	88	925

Stacking Classifier

		Predicted Label	
		World News	The Onion
True Label	World News	879	118
	The Onion	96	917

Sources Referenced

- [1] <https://www.reddit.com/r/worldnews/>
- [2] <https://www.reddit.com/r/TheOnion/>
- [3] <https://www.theonion.com/supreme-court-overturms-car-1819564540>