

Sarcasm Detection Using Natural Language Processing

By Adam Pell

Outline

Business Problem

Data Understanding

Methodology

Results

Recommendations

Sarcasm is Hard



Tone and context dependent

Sarcasm relies heavily on tone of voice and context, which can be difficult to pick up in written text.



Requires understanding nuance

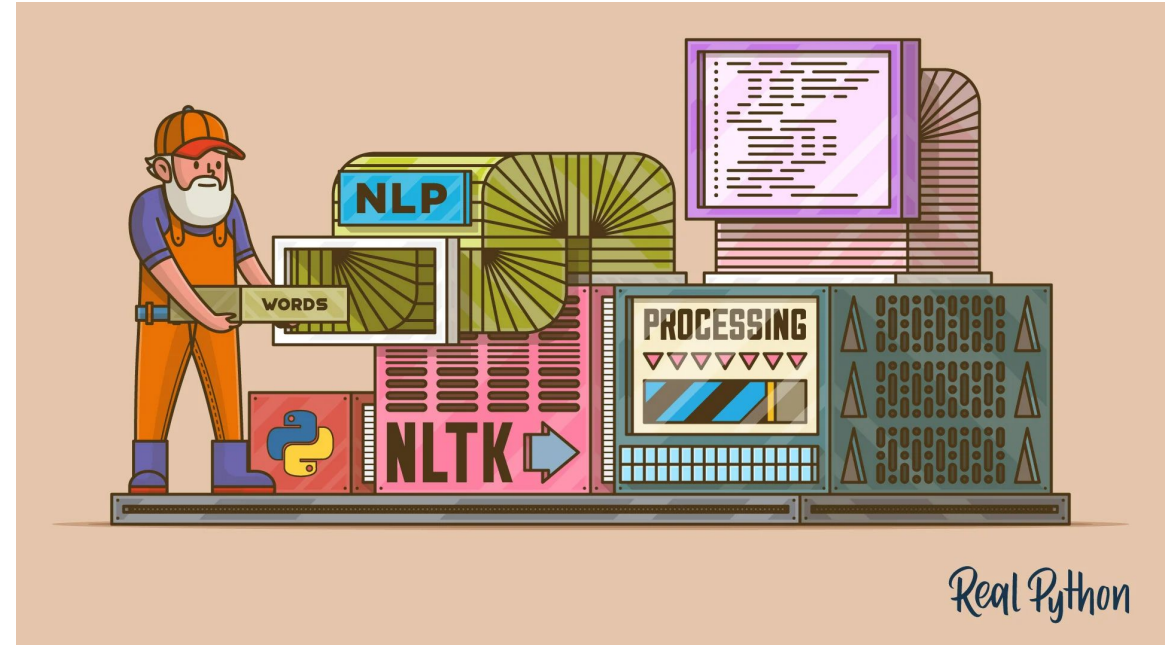
Detecting sarcasm requires understanding subtle nuances in language and meaning.



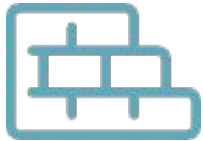
Difficult to detect in written text

Without vocal cues, sarcasm can be very difficult to identify in written communication.

The Solution: Natural Language Processing



Applications



Bullying Detection

Differentiate harmful behavior, resulting in bans and suspensions



Inappropriate Behavior

Detect genuine threats or concerning language



Algorithm Enhancement

Deliver similar content to people who interact heavily with sarcastic posts

Understanding sarcasm opens up a world of possibilities

Dataset Overview

- 1m+ entries
- Comments and parent posts
- Labeled for sarcasm
- Balanced



Methodology



Supervised Learning



Different Models



Target Metric: **Recall**

Model Performance Compared

Model	Recall (Class 0)	Recall (Class 1)	Accuracy	Time to Run
Naive Bayes	74%	55%	64%	>1s
XGBoost	94%	21%	58%	5min.
Neural Network	75%	57%	66%	21min.
Recurrent Neural Network	91%	16%	54%	31min.

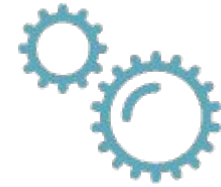
Results: Final Model Key Trends



Most Consistent
Least variation in recall

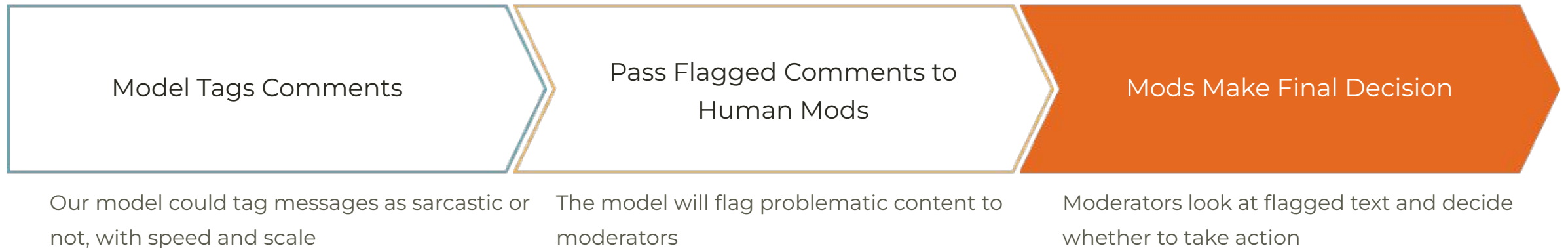


Best Sarcasm Detection
Achieves the highest recall for Class 1
(sarcasm)



Further Training Needed
Sarcasm detection still proves difficult

Recommendation: A Hybrid Approach



Next Steps

- Data quality

Better labeling practices

- Transfer learning

Fine-tune using large, pre-trained model

- Optimize models

Longer, more intensive training and experimentation

Thank You!

Contact Information

- **Email:** apell7591@gmail.com
- **Phone:** 917-434-6615
- **Github:** [apell74](https://github.com/apell74)
- **LinkedIn:** [Adam Pell](#)

Relevant Links and Sources

- [Project Repository](#)
- [Google Colab Notebook](#)
- [Kaggle Dataset](#)
- [Further Reading](#)