

# Spam Detection on YouTube

Mustafa Kolanci

# Agenda



1 Business Understanding

2 Dataset Overview

3 Methods

4 Findings & Recommendations

5 Future Research

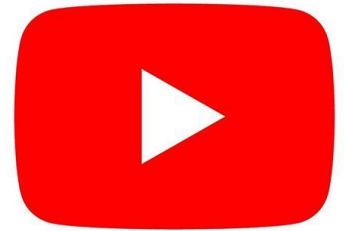
# Spam detection

93.6% accuracy

# Business Understanding

# Improve User Experience

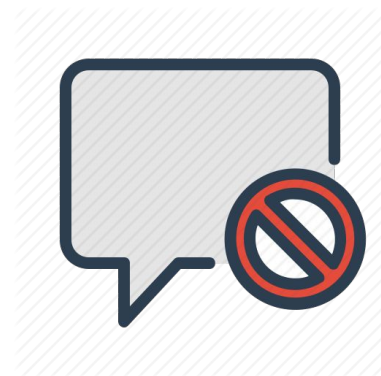
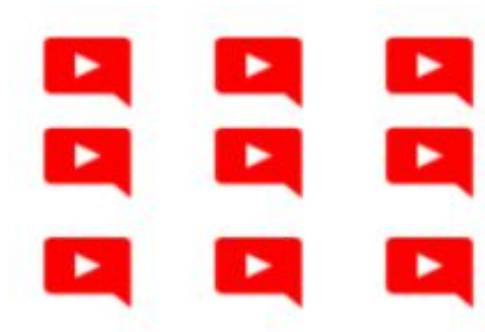
Identifying spam comments on YouTube



Data

# Dataset

- 2013-2015
  - PSY, Katy Perry, Shakira, Eminem, LMFAO
- ~2,000 comments
- Comments: # spam = # non-spam



# Methods



# Methods

- Eliminate
  - Html tags
  - Url's
  - Commonly used words
- Accuracy



# Key Findings

## Most Important Words

- Channel
- Check
- Like
- Please
- Subscribe

## Most Important Bigrams

- Check channel
- Check video
- Check new

# Recommendations

# Recommendations

- Only music videos
- Flag comments



# Future Research

# Areas of Future Research

- Non-English comments
- Identify spam accounts
- Expand to non-music videos

What's Next?



# Thank you!



**Mustafa Kolanci**



`mfurkankolanci`



`mfurkankolanci@gmail.com`