

NLP on Twitch Chat to build Videos Highlights

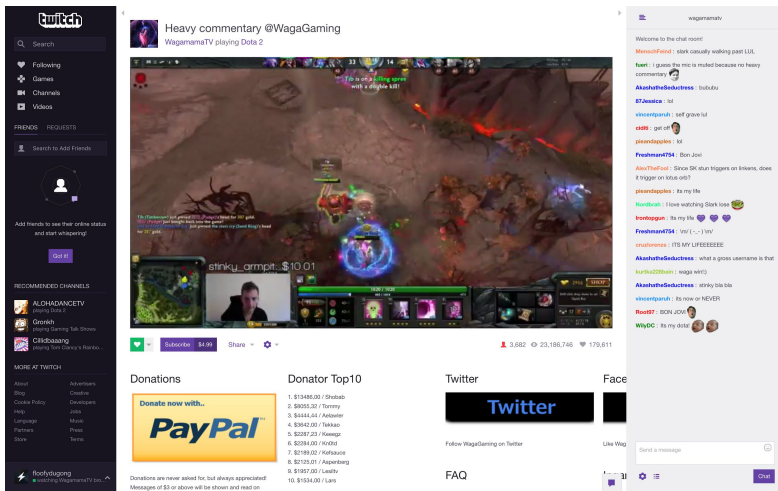


John Huynh



Methodology

1. Scraped Twitch data from **150 streamers** (their broadcasts and past highlight videos)
2. **1.7 Million texts** Twitch chat logs from past recordings of streams related to Dota
3. Attempted to learn “twitch” language and **feature engineer**
4. Identified topics around using **TF-IDF** and **latent semantic analysis** for one video to as an example
5. Generated links that would enable users to quickly scrub through the video



Twitch

“Twitch is the world’s leading social video platform and community for gamers”



DOTA 2

“Dota 2 is a free-to-play multiplayer online battle arena (MOBA) video game developed and published by Valve Corporation”

Twitch Chat Example

[illegible]

What People Are Saying

**twitch chat =
cancer of the
universe**

Twitch User

**HEY IM FROM
ASIA AND AM AN
ASIAN IM NOT
RACIST 4Head**

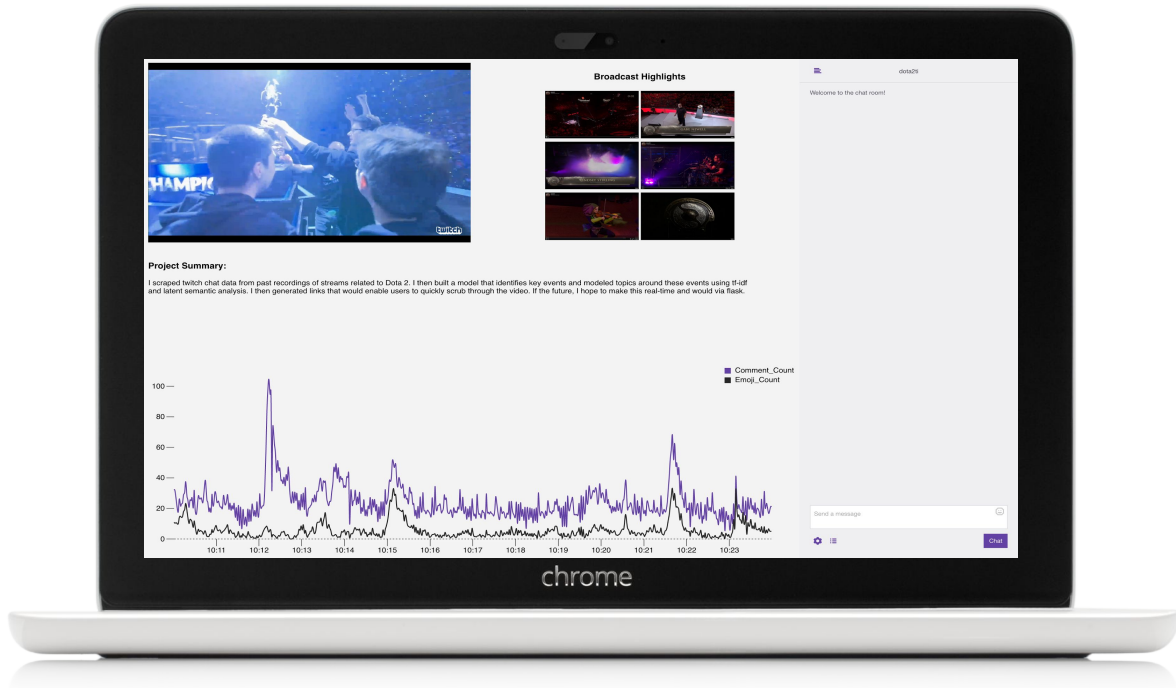
Twitch User

**Reading twitch
chat reversed my
Lasik**

Metis Student

Demo!

Web App Demo with D3.js visualization



Lessons Learned

Takeaways

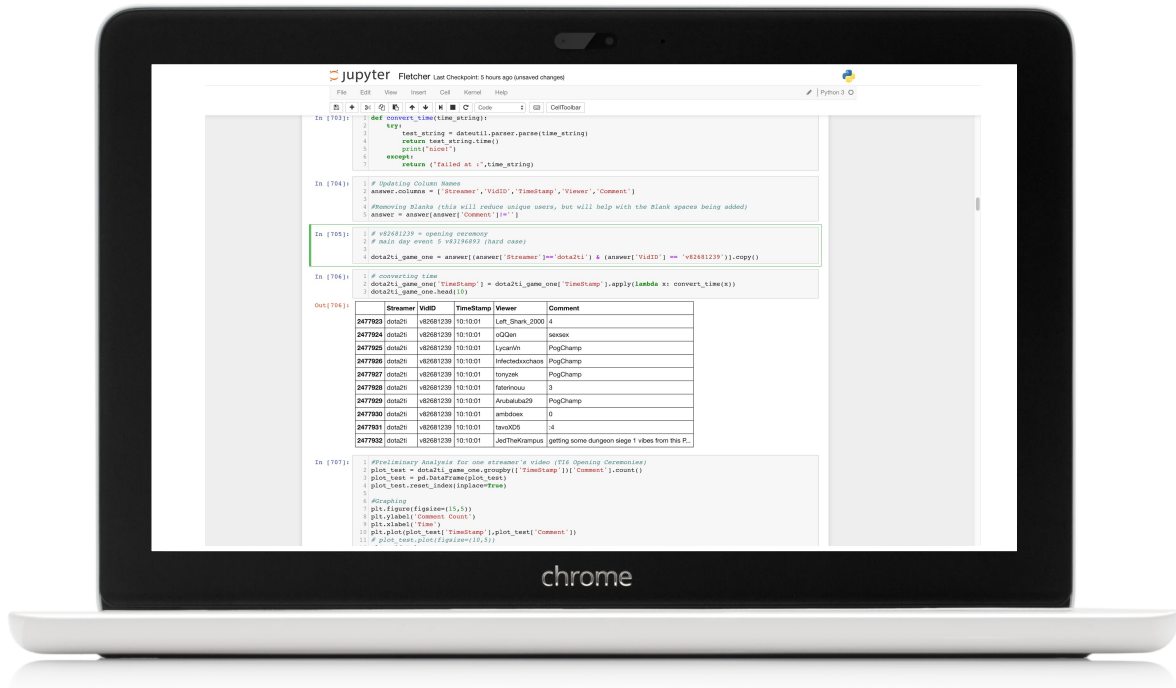
- Twitch chat is interpretable to some degree
- Review chat and learn the “language”
 - Stop Words
 - Determine what a corpus means
- Find a balance between explained variance ratio and word “strength” per component
- Continue to feature engineer and tinker with the word vectorizers

Improvements for Next Time

- Dynamically create content (links, thumbnails, etc.)
- Sync the Video and D3
- Integrate replay Chat
- Pipeline the steps better
- Create a ratio of users/messages
- Flask More D3

Jupyter Notebook

Walkthrough of Jupyter Notebook



Credits:
[Twitch](#)
[Wikipedia](#)
[Mike Bostock](#)
[freaktechnik](#)

