DISCOVERY CHALLENGE

# ChAT

## Chat Analytics for Twitch

Konstantin Kobs Martin Potthast Matti Wiegmann Albin Zehe Benno Stein Andreas Hotho

# Introductory Round

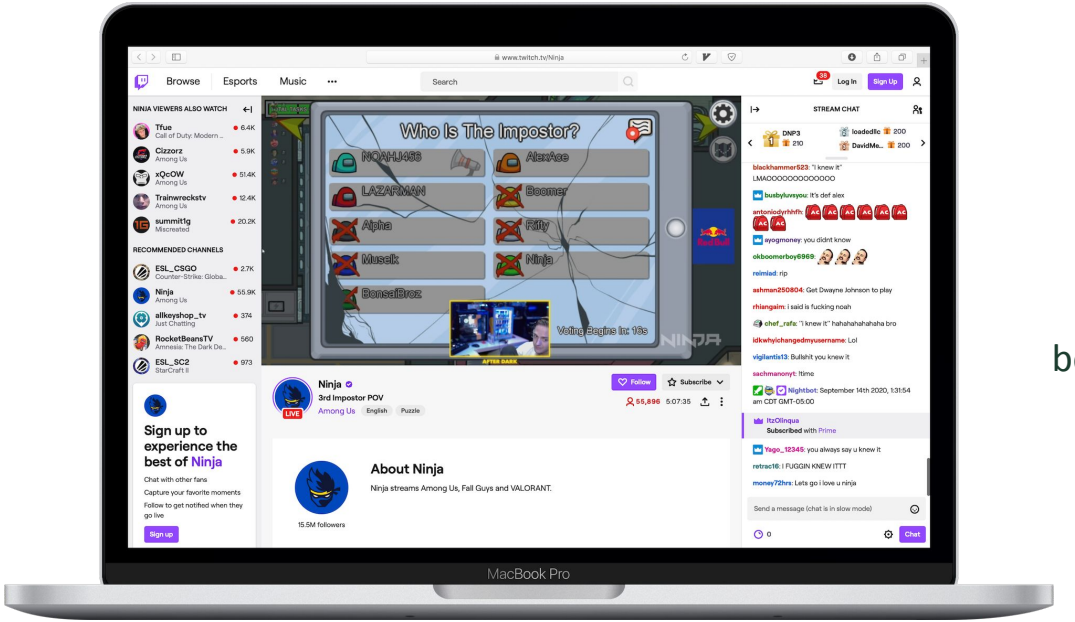
Just say your **name**, **affiliation**, and **why you attend** this session!

Twitch.tv

# Twitch: Game Streaming Platform

launched

2011




2014

bought by Amazon

# Twitch: Key Performance Indicators

**2** million concurrent viewers

**90 000** concurrent streams

**> 30** languages

**> 1.5 billion** hours watched per month (since corona)

**1** million unique daily streamers

**7 million** active streamers

**1 billion** public Twitch chat comments per month

https://twitchtracker.com/

# Twitch: Business Model

| 4.99$/month | 9.99$/month | 24.99$/month |
| --- | --- | --- |
| **Benefits** | **Benefits** | **Benefits** |
| ad-free viewing of the channel | everything from 4.99$/month | everything from 9.99$/month |
| subscriber-exclusive chat rooms | more emotes | even more emotes |
| subscriber-only emotes | | |
| subscriber badge | | |

**Top 20 channels** generate nearly **800 000$/month** through subscriptions alone[1]

Revenue through advertisements before and during streams, sponsorings, ...

50% for the streamer, 50% for Twitch

[1] estimated with numbers from https://twitchtracker.com/subscribers

# Twitch: Research Subject

**Research Questions**

- When and why are users watching?
- What games do users like to watch?
- How do chat users interact with each other?
- Do users like what they see?
- Do subscribers behave differently than non-subscribers in chat?
- ...

**Research Fields**

- Natural Language Processing
- Image Processing
- Audio Processing
- Recommender Systems
- Graph Analysis
- Social Studies
- ...

# ChAT Discovery Challenge

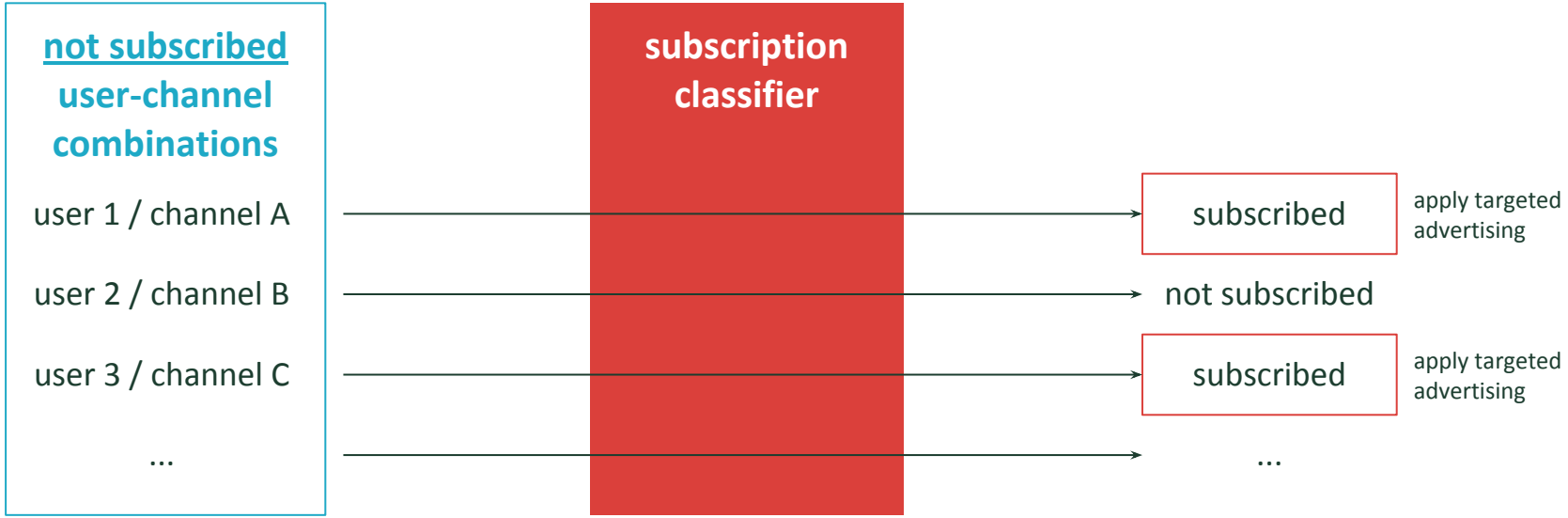
## Question

Do subscribers behave differently than non-subscribers in chat?

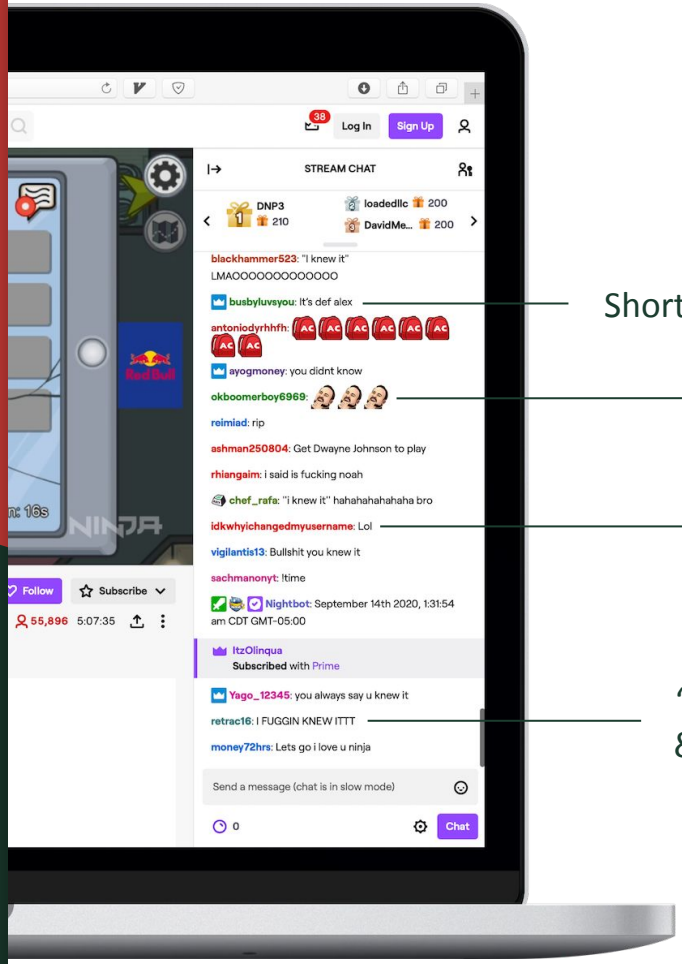
## Task

“Build a **binary classifier** that predicts the **subscription status** of users in Twitch channels **given their chat messages**”

# Application




**Follow-up Question:** Do **unsubscribed** users who are **misclassified** as subscribed to a channel make good marketing targets?




Short messages

Emotes

Slang

"Shouting" & mistakes

# Twitch, another Language

- Emotes are crucial to Twitch's language
- Fairly different from common English
- Emotes provide a way to express opinions and emotions

| Rank | Twitch | Twitch (no emotes) | English | Twitter |
|---|---|---|---|---|
| 1 | LUL | **like** | new | tinyurl.com |
| 2 | Kappa | **get** | home | new |
| 3 | <3 | lol | us | **like** |
| 4 | PogChamp | u | page | **good** |
| 5 | **like** | **good** | search | **get** |
| 6 | **get** | **2** | free | **time** |
| 7 | lol | 1 | **one** | day |
| 8 | :D | game | information | **one** |
| 9 | Kreygasm | stream | **time** | twitter |
| 10 | Clap | **got** | site | going |
| 11 | u | **one** | may | **go** |
| 12 | **good** | **go** | news | rt |
| 13 | :) | play | use | **know** |
| 14 | **2** | xD | see | today |
| 15 | 1 | 3 | contact | love |
| 16 | game | **know** | business | work |
| 17 | HeyGuys | **time** | web | **got** |
| 18 | BibleThump | **think** | also | **2** |
| 19 | stream | see | help | **back** |
| 20 | **got** | **back** | **get** | **think** |

Kobs, Konstantin, et al. "Emote-Controlled: Obtaining Implicit Viewer Feedback Through Emote-Based Sentiment Analysis on Comments of Popular Twitch.tv Channels." *ACM Transactions on Social Computing* 3.2 (2020): 1-34.

# Emote Intensification

- Word2Vec trained on Twitch chat messages
- Intensification of emotes possible

→ Emotes capture meaning


Kobs, Konstantin, et al. "Emote-Controlled: Obtaining Implicit Viewer Feedback Through Emote-Based Sentiment Analysis on Comments of Popular Twitch.tv Channels." *ACM Transactions on Social Computing* 3.2 (2020): 1-34.

# Emote Lexicon

Build a sentiment classifier for Twitch chat messages

→ Able to get real-time sentiment of a stream's audience

| Emote \ Sentiment | Negative | Neutral | Positive | Unknown/NA |
|---|---|---|---|---|
| FeelsBadMan | 71 | 17 | 19 | 1 |
| FeelsGoodMan | 1 | 7 | 98 | 2 |
| LUL | 11 | 23 | 72 | 2 |
| OMEGALUL | 17 | 26 | 62 | 3 |
| PogChamp | 1 | 3 | 101 | 3 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| Jebaited | 25 | 27 | 37 | 19 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| mcaT | 10 | 34 | 12 | 52 |
| forsenPls | 13 | 26 | 17 | 52 |
| PepoDance | 9 | 26 | 20 | 53 |
| RedCoat | 5 | 46 | 4 | 53 |
| jinnytHype | 7 | 31 | 17 | 53 |

Kobs, Konstantin, et al. "Emote-Controlled: Obtaining Implicit Viewer Feedback Through Emote-Based Sentiment Analysis on Comments of Popular Twitch.tv Channels." *ACM Transactions on Social Computing* 3.2 (2020): 1-34.

# Twitch: Research Subject

**Research Questions**

- When and why are users watching?
- What games do users like to watch?
- How do chat users interact with each other?
- **Do users like what they see?**
- Do subscribers behave differently than non-subscribers in chat?
- ...

**Research Fields**

- **Natural Language Processing**
- Image Processing
- Audio Processing
- Recommender Systems
- Graph Analysis
- Social Studies
- ...

# Twitch: Research Subject

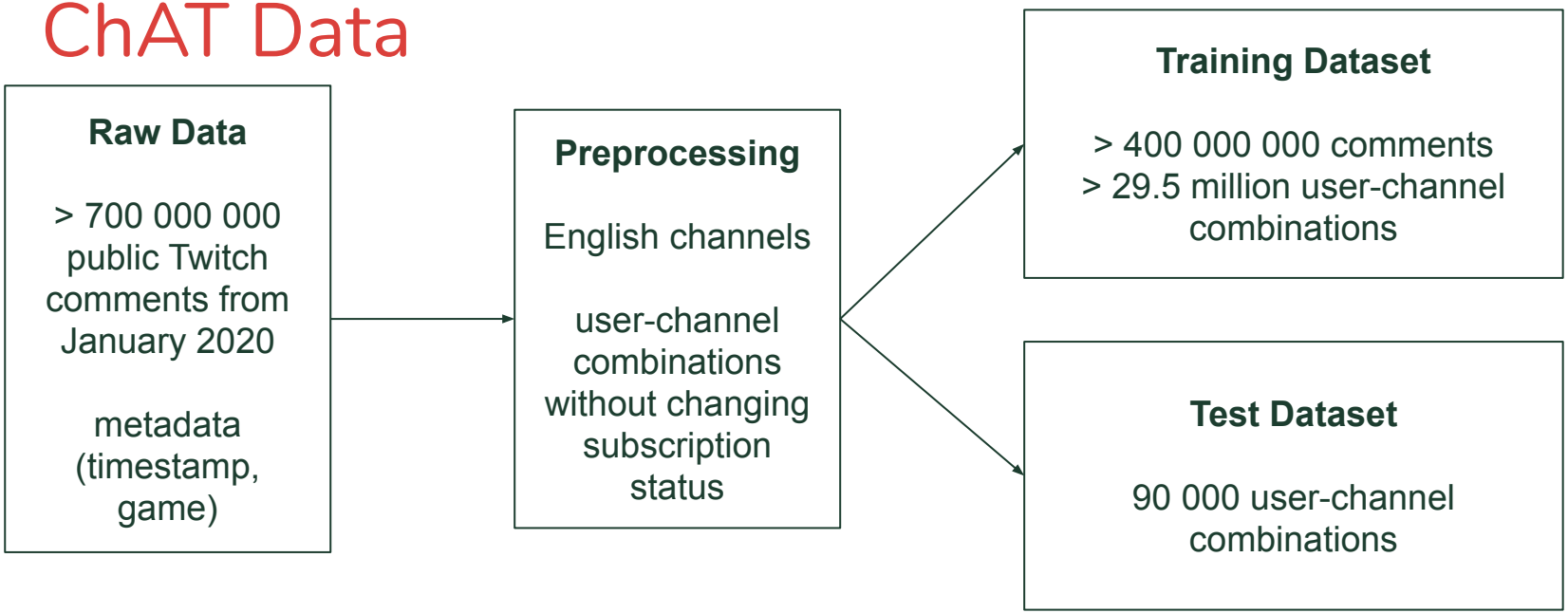
**Research Questions**

- When and why are users watching?
- What games do users like to watch?
- How do chat users interact with each other?
- Do users like what they see?
- **Do subscribers behave differently than non-subscribers in chat?**
- ...

**Research Fields**

- Natural Language Processing
- Image Processing
- Audio Processing
- Recommender Systems
- Graph Analysis
- Social Studies
- ...

# ChAT Data




→ download both datasets on the challenge's website

# Testset Sampling

- Categorize users and channels into **low** (25%) / **normal** (50%) / **high** (25%) activity groups
- 10 000 user-channel combinations for each activity group combination ⟶ 90 000
  - **low** user — **low** channel
  - **low** user — **normal** channel
  - …
  - **high** user — **high** channel
- For **50% of test users**, remove all comments from other channels in the training dataset

⟶ Analyze how models handle **different user and channel activities** and **new users**

# Training Dataset

- ~150 000 channels
- ~8 million users
- **29.5 million user-channel combinations** that **had at least one comment**
- 8.02% **subscribed**, 91.98% **not subscribed**
- On average, users comment in **3.73 channels** and are subscribed to **1.5 channels**
- Comments per user in channel:
  - **subscribed**: 55.66
  - **not subscribed**: 43.08
- Number of comments per channel: 2802.61

# Evaluation

**Performance Measure**

Precision
Recall
**$F_1$ score**

**Baseline**

Sampling randomly from the training distribution:

8.02% **subscribed** ↔ 91.98% **not subscribed**

software submission process over **TIRA.io**  tira

# Statistics

3 months to complete the task

23 team registrations

4 model submissions over TIRA.io

3 descriptive papers

# Results

| Rank | Team | Precision | Recall | $F_1$ | Runtime |
|---|---|---|---|---|---|
| 1 |  VoyTECH | 0.2796 | 0.4446 | 0.3433 | 00:07:39 |
| 2 |  CoolStoryBob | 0.1904 | 0.4341 | 0.2647 | 00:05:34 |
| 3 |  ItsBoshyTime | 0.4808 | 0.1775 | 0.2593 | 00:00:19 |
| 4 |  StinkyCheese | 0.0817 | 0.5487 | 0.1422 | 00:13:06 |
| | Random Baseline | 0.0689 | 0.0802 | 0.0741 | |

# Team ItsBoshyTime 

- Exploit some shortcomings of the dataset:
  **If users are subscribed to a channel, they can use subscriber-exclusive emotes**
- Idea
  - **Extract subscriber-only emotes** from training dataset using a heuristic
  - For a new user-channel combination: **Does the user uses subscriber-only emotes from this channel?**
    - If yes: predict **subscribed**
    - If no: predict **not subscribed**

→ **Disqualified**

# Results

**CoolStoryBob**

2

Marvin Gärtner
Andreas Theissler
Marc Fernandes

**voyTECH**

1

Immanuel Bayer
Anastasios Zouzias

**StinkyCheese**

3

Túlio Corrêa Loures
Gustavo Lúcius Fernandes
Fernanda G. Araújo
Karen S. Martins
Pedro O. S. Vaz de Melo

# Agenda for this Session

**15:00 – 15:15:** Welcome and overview talk by the organizers

*"Towards Predicting the Subscription Status of Twitch.tv Users"*

Konstantin Kobs, Martin Potthast, Matti Wiegmann, Albin Zehe, Benno Stein, Andreas Hotho

**15:15 – 15:45:** Talk by team voyTECH and Q&A

*"User Activity Modeling with Boosting Trees"*

Immanuel Bayer, Anastasios Zouzias

**15:45 – 16:15:** Talk by team CoolStoryBob and Q&A

*"Detecting Potential Subscribers on Twitch: A Text Mining Approach with XGBoost"*

Marvin Gärtner, Andreas Theissler, Marc Fernandes

**16:15 – 16:45:** Talk by team StinkyCheese and Q&A

*"Chat-Based Model for Subscription Classification"*

Túlio Corrêa Loures, Gustavo Lúcius Fernandes, Fernanda G. Araújo, Karen S. Martins, Pedro O. S. Vaz-de-Melo

ChAT website

# Team voyTECH

*“User Activity Modeling with Boosting Trees”*

Immanuel Bayer, Anastasios Zouzias

# Team CoolStoryBob

*“Detecting Potential Subscribers on Twitch: A Text Mining Approach with XGBoost”*

Marvin Gärtner, Andreas Theissler, Marc Fernandes

# Team StinkyCheese

*“Chat-Based Model for Subscription Classification”*

Túlio Corrêa Loures, Gustavo Lúcius Fernandes, Fernanda G. Araújo,
Karen S. Martins, Pedro O. S. Vaz-de-Melo