Identifying the Best Streaming Strategy That Gain Most Followers On Twitch

By Ha Le, Junming Wu

Background

- Twitch is the most popular streaming platform made for competitive streaming and eSport. Currently, there are 41.5 million in the state, and it is estimated to reach 51.6 million by 2024.

- Streamers are try to build a large community under their name to attract sponsorship and partnership, which could be another source of income for full-time streamers and eSport players.

- We are trying to fnd the strategies for streamers to gain more followers/viewers. Given their streaming activities we want to predict whether they will do well over the course of one week.

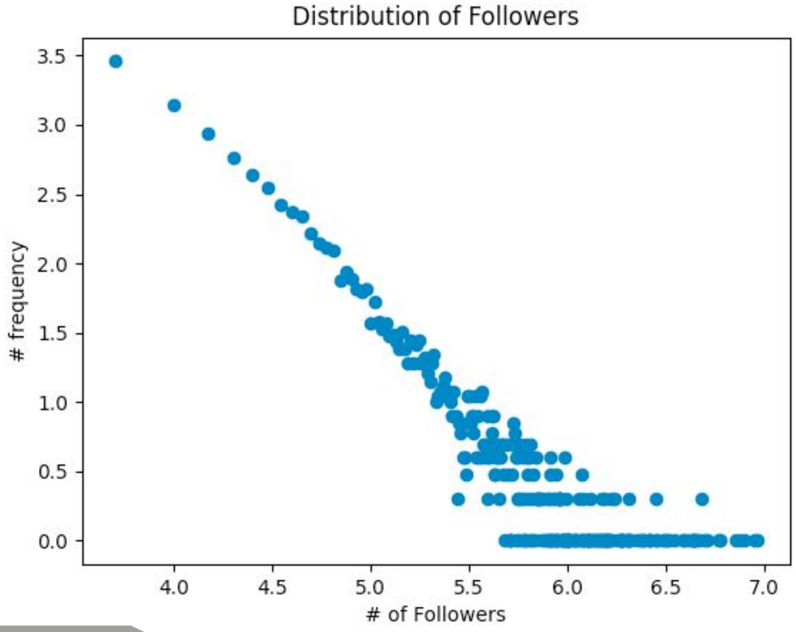
Crawling Process

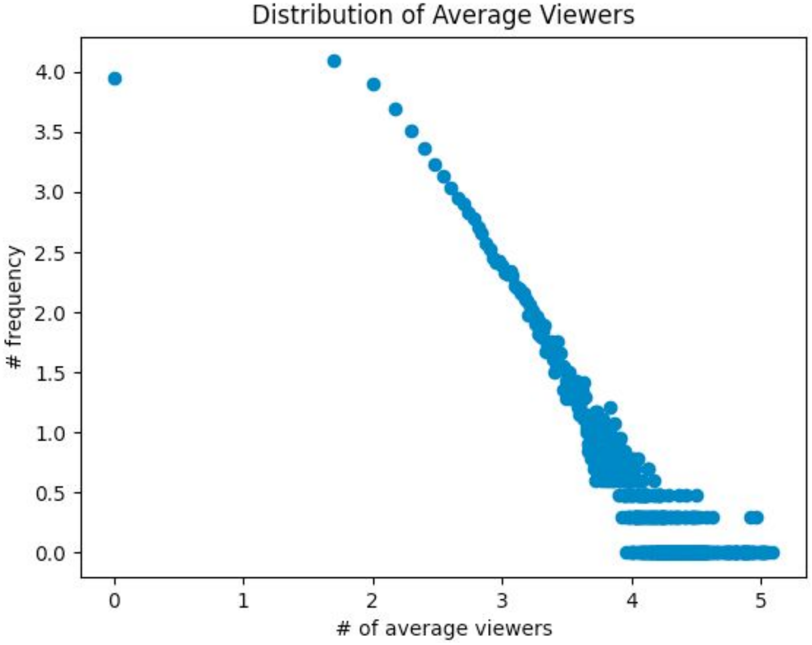
Step 1: Generate 10000 active users
Step 2: monitor 4000 users per week, doing the experiment over 4 weeks

1. Go through every user and get their statistics every 30 minutes
2. If the user is inactive, grasp information about their number of followers
3. If the user is streaming, get the information about the stream
   I. Inactive: current_time, user_id, # followers
   II. Streaming: current_time, user_id, # followers, #viewers, game_id, game_name, start_at, language

Step 3: Gathering information about archive videos and clips posted after the stream

Power Law Distribution


Distribution of Followers


Distribution of Average Viewers

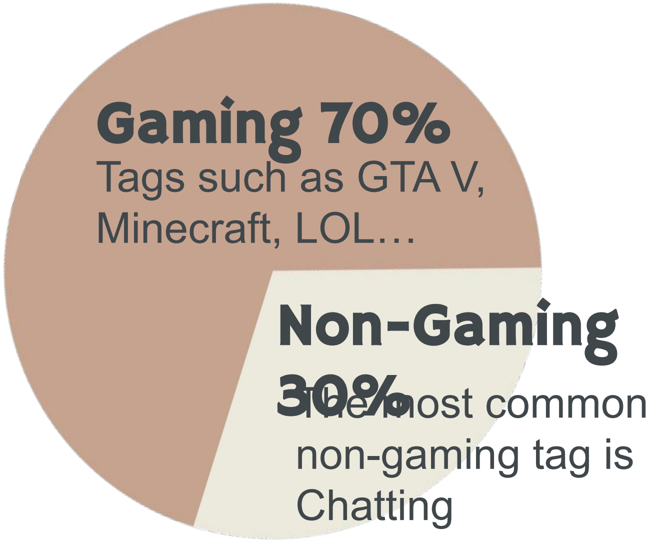
Isolating the Outliners

To identify an effective strategies to gain more followers during streaming periods, we will isolate streamers who gained more followers compared to other streamers at the same popularity.

I. Divide streamers' initial popularity range into 20 intervals: 0-1000, 1000-2000, 2000-3000, …
II. For each interval, choose 5% streamers who gains the most amount of followers over the course of one week.
III. Look at what they have in common to devise a strategies
IV. Identify streamers who have been adopting the strategies, measure their performance and compare it to the rest of the dataset

We were able to get 197 outliners from 5731 users with the initial followers' number between 0 and 20000.

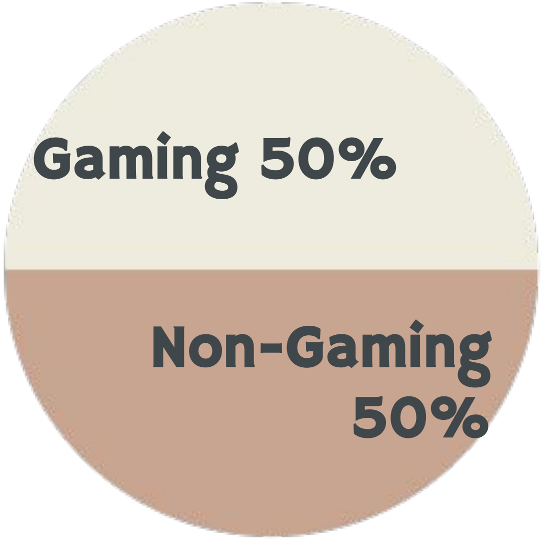
Content Distribution



Tags such as GTA V, Minecraft, LOL…

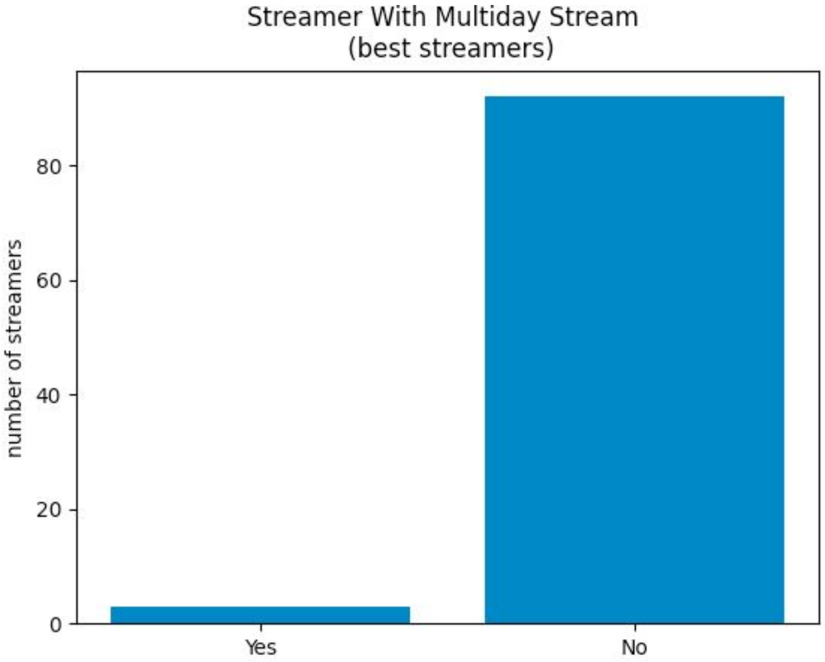
The most common non-gaming tag is Chatting

Entire Dataset



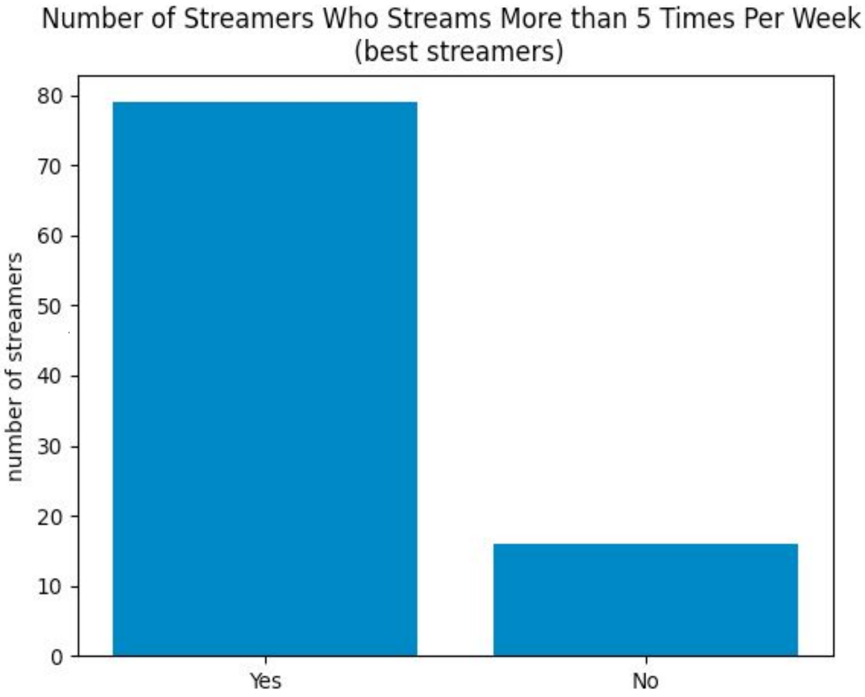
Outliners

Length and Frequency of Streaming



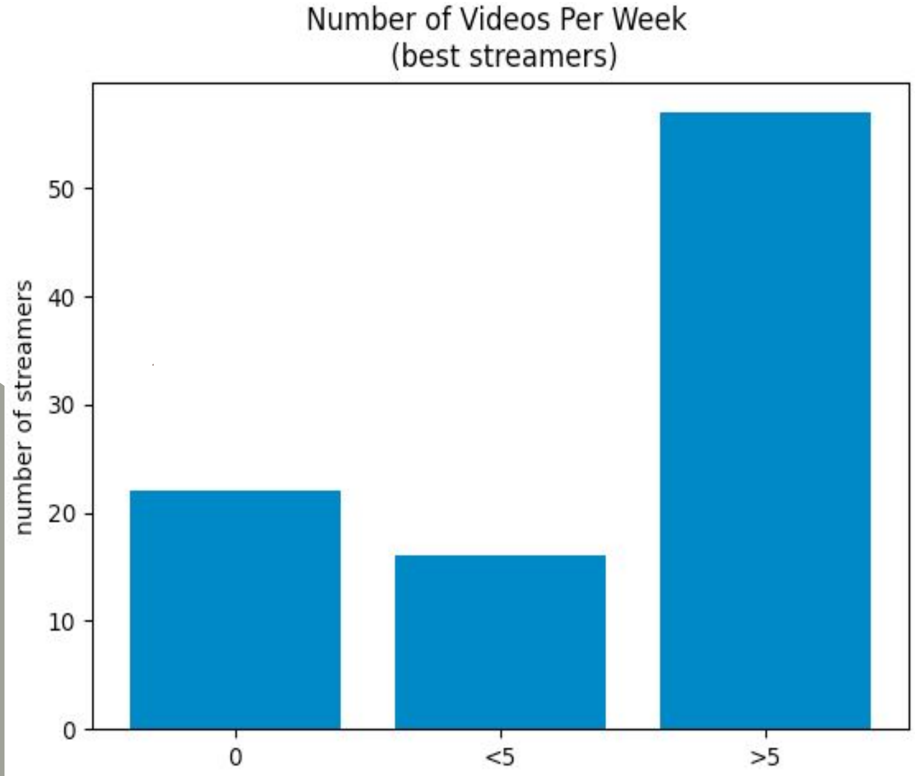
Most outliners do not stream over 24 hours.

Length and Frequency of Streaming



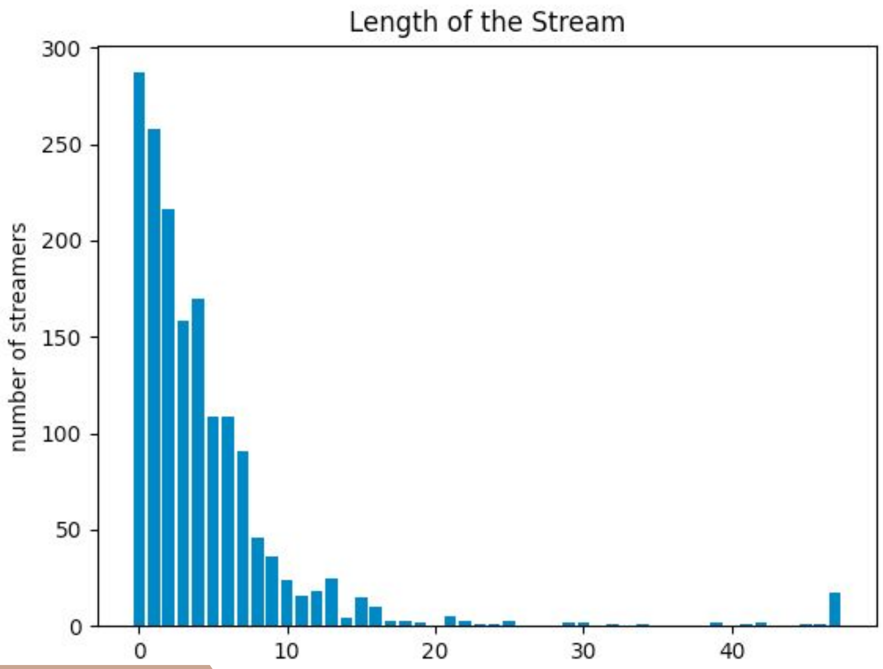
Most outliners stream more than 5 times per week.

Length and Frequency of Streaming

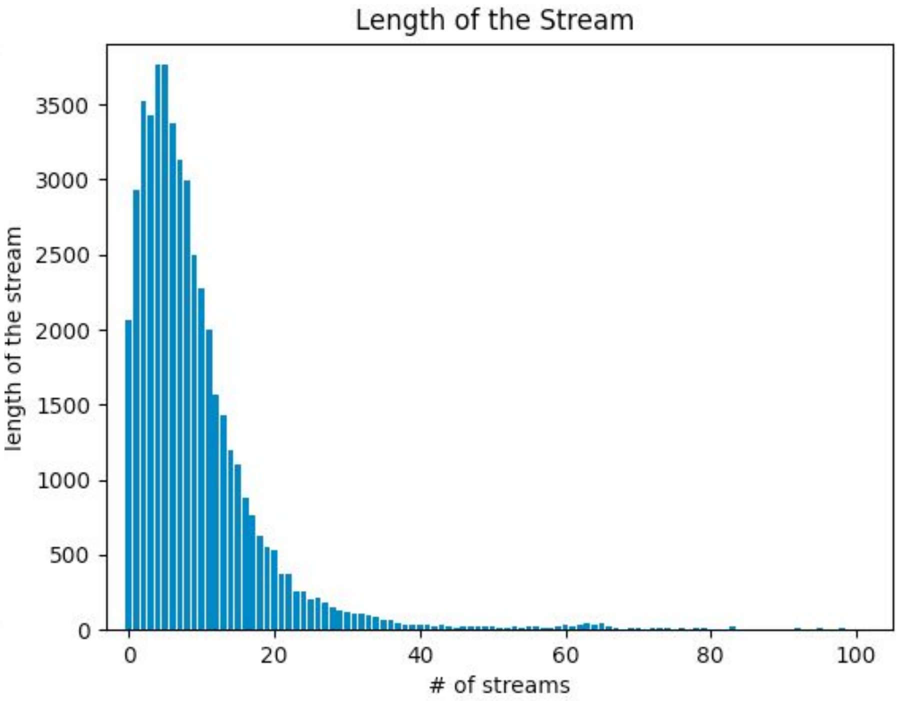


Most outliners post over 5 videos per week.

Length of the Stream

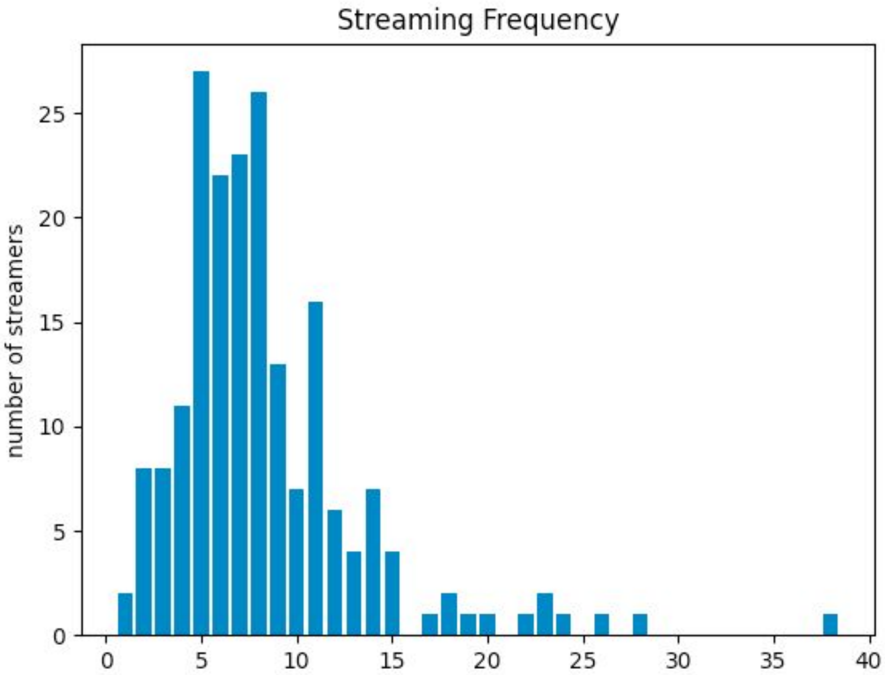


Outliners

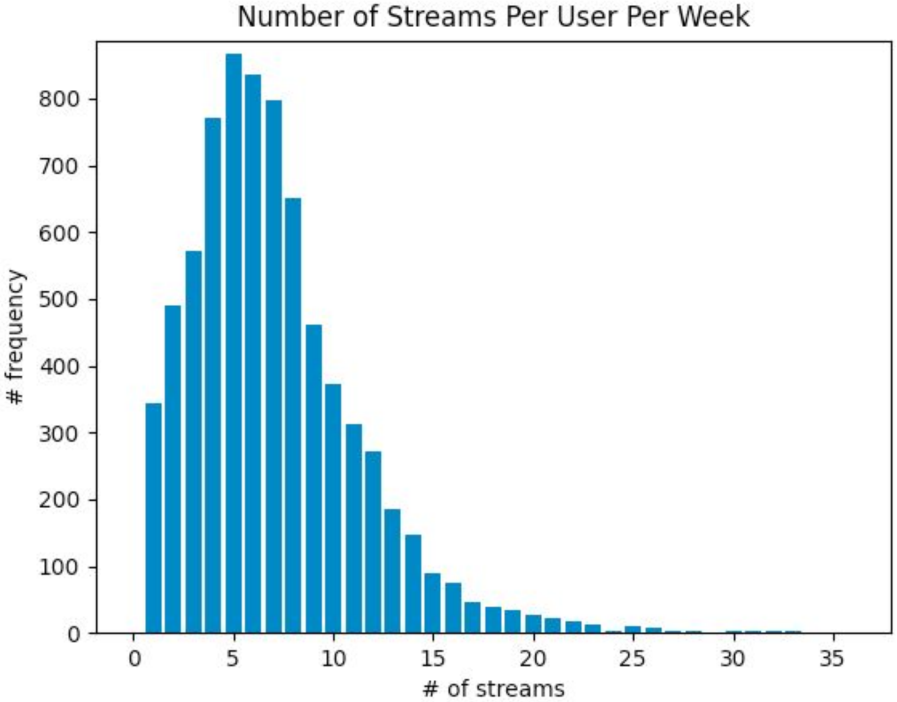


Entire Dataset

Streaming Frequency



Outliners



Entire Dataset

Strategies

- Stream at least 5 times per week
- Each stream lasts less than 12 hours
- Mixed non-gaming content with gaming content
- Stream more than 40 hours per week

Comparison

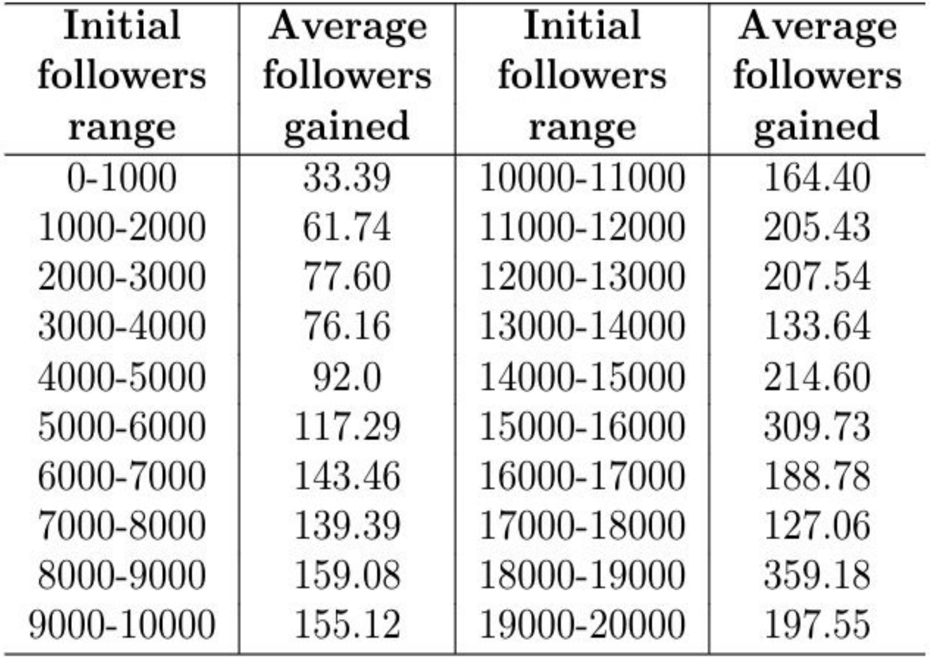
	Entire Dataset	Streamers w this strategy
Average Follower Gained	101	342
Median Follower Gained	29	67
Average Viewers per Hour	101	101
# streamers who lost followers	321/5731	1/51

Predicting the Growth of a Channel

- Classify users into classes based on their weekly growth rate
- Focus on streamers with initial number of followers between 0-20k
- Use a random forest classifier
- Random forest classification is an ensemble based learning algorithm. It involves the usage of multiple decision trees when training, and majority votes in the prediction process.

Predicting the Growth of a Channel

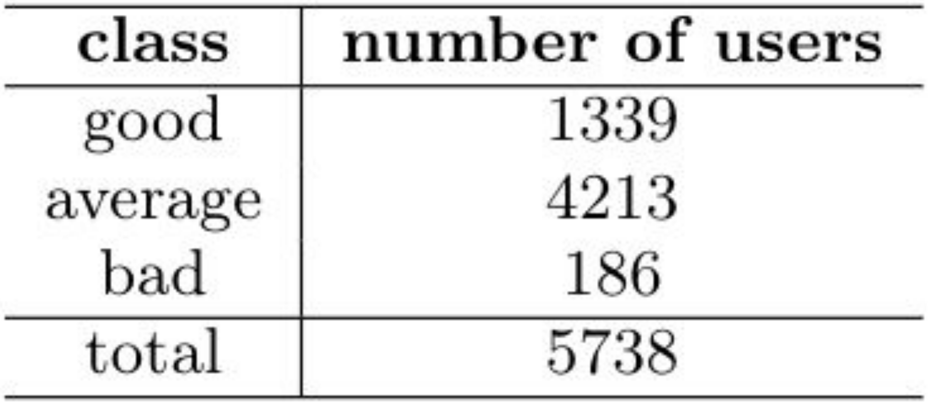
- Divide streamers into 20 different bucket based on their initial number of followers: 0-1000, 1000-2000, …, 19000-20000
- Divide the streamers into 2 sets of data: the training set (80%) and the testing set (20%)
- Calculate the average followers gained of each bucket in the training set.



Initial followers range	Average followers gained	Initial followers range	Average followers gained
0-1000	33.39	10000-11000	164.40
1000-2000	61.74	11000-12000	205.43
2000-3000	77.60	12000-13000	207.54
3000-4000	76.16	13000-14000	133.64
4000-5000	92.0	14000-15000	214.60
5000-6000	117.29	15000-16000	309.73
6000-7000	143.46	16000-17000	188.78
7000-8000	139.39	17000-18000	127.06
8000-9000	159.08	18000-19000	359.18
9000-10000	155.12	19000-20000	197.55

Classifying the Users

- Based on the average followers gained of each bucket, classify the users into 3 different groups:
- *Bad* streamers: streamers who lost followers.
- *Good* streamers: streamers who gained more followers then the average gained followers of their bucket.
- The rest is label *average* streamers



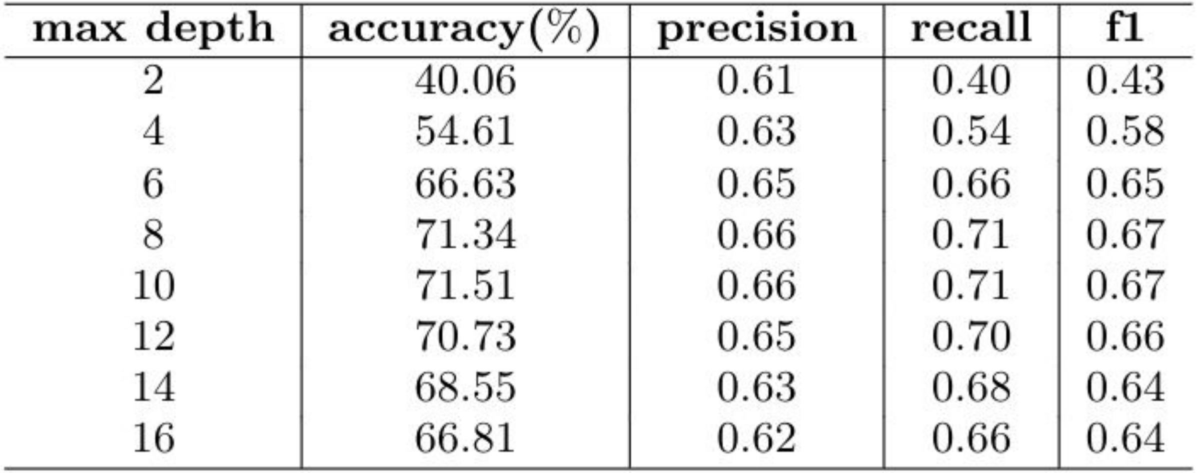
class	number of users
good	1339
average	4213
bad	186
total	5738

Predicting the Growth of a Channel

- Since the class’s sizes are unbalanced, we used a synthesis minority oversampling technique (SMOTE) on the training set to improve the accuracy of the model
- Next, we feed the data to the model, with the following learning parameters:
  - Initial number of followers and their bucket
  - Number of streams last less than 5 hours (in one week)
  - Number of streams last between 5-10 hours (in one week)
  - Number of streams last more than 10 hours (in one week)
  - Whether they stream non-gaming content
  - Whether they stream popular games (top 5 most popular game tag on Twitch)
- The model will return the predicted classification of the users: 1=good, 0=average, -1=bad

Predicting the Growth of a Channel

- After tuning the parameters, we figure out our model will work best when we set the maximum depth of the trees = 10, with the accuracy rate of 71.55%



max depth	accuracy(%)	precision	recall	f1
2	40.06	0.61	0.40	0.43
4	54.61	0.63	0.54	0.58
6	66.63	0.65	0.66	0.65
8	71.34	0.66	0.71	0.67
10	71.51	0.66	0.71	0.67
12	70.73	0.65	0.70	0.66
14	68.55	0.63	0.68	0.64
16	66.81	0.62	0.66	0.64

Thanks