Youtube Fame Predictor

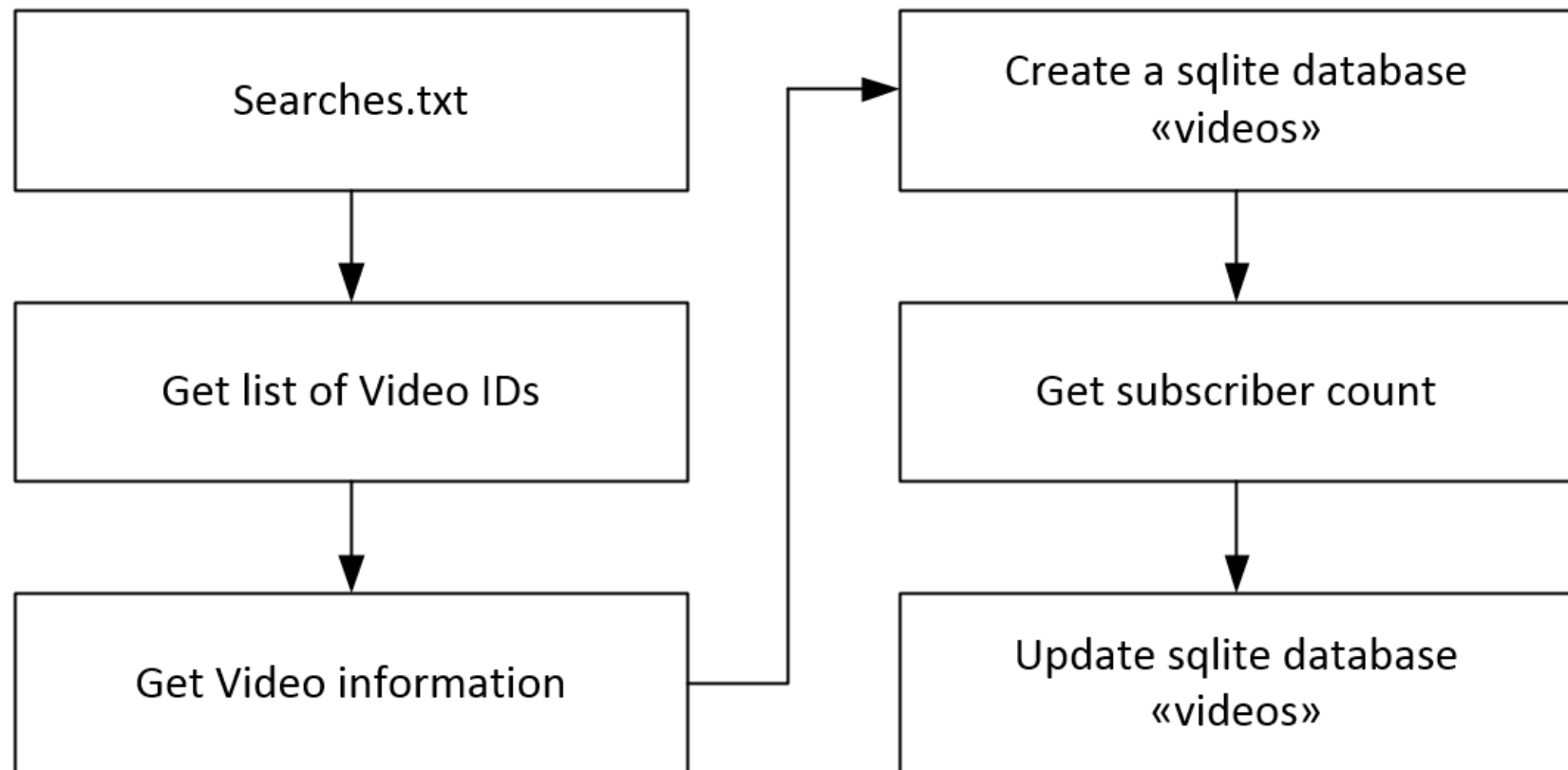## Network Tour of Data Science

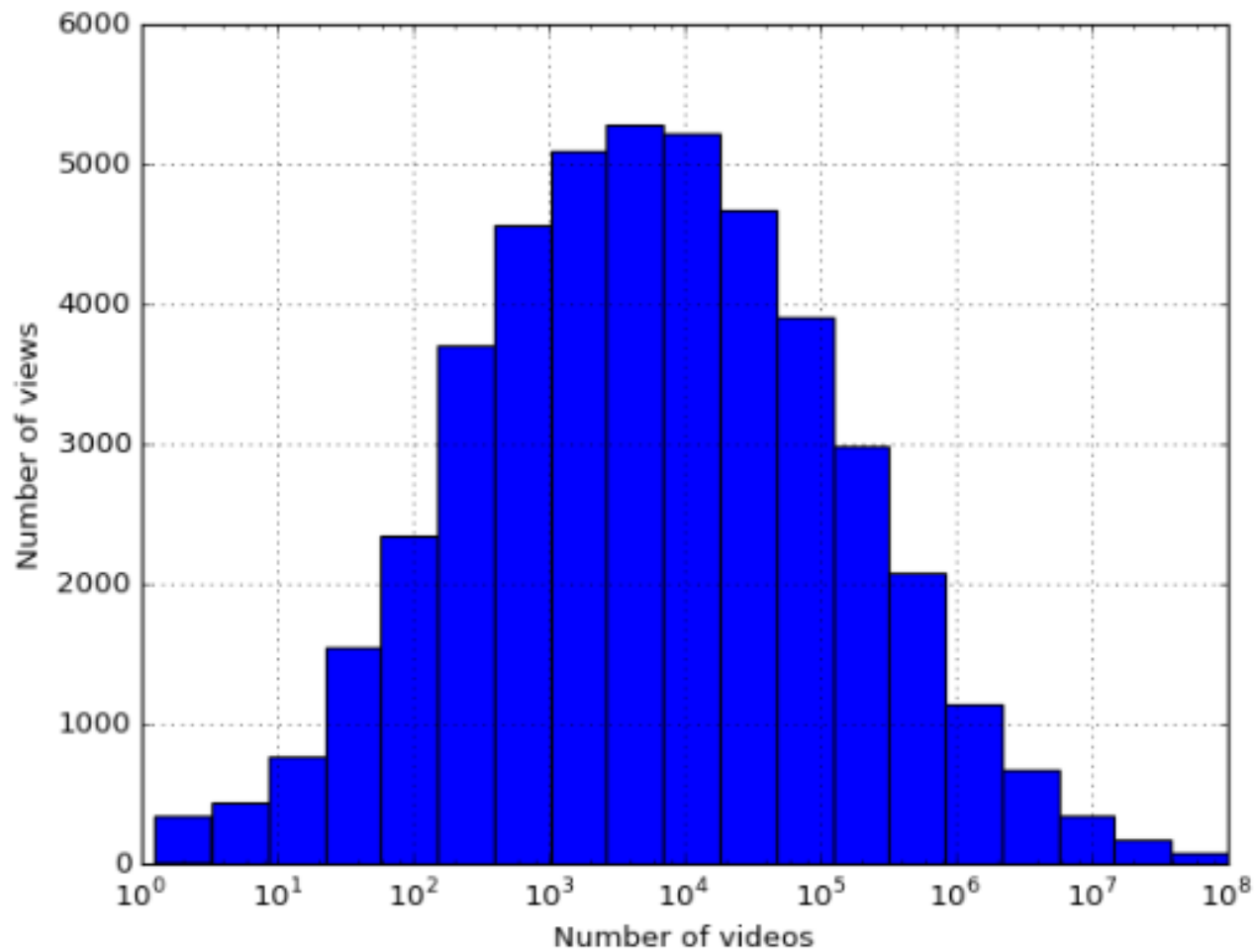Cyrille Rolland

Benoît Steinmann

Introduction

- Goal : predict the number of views of a Youtube videos with:
  - Thumbnail of the video
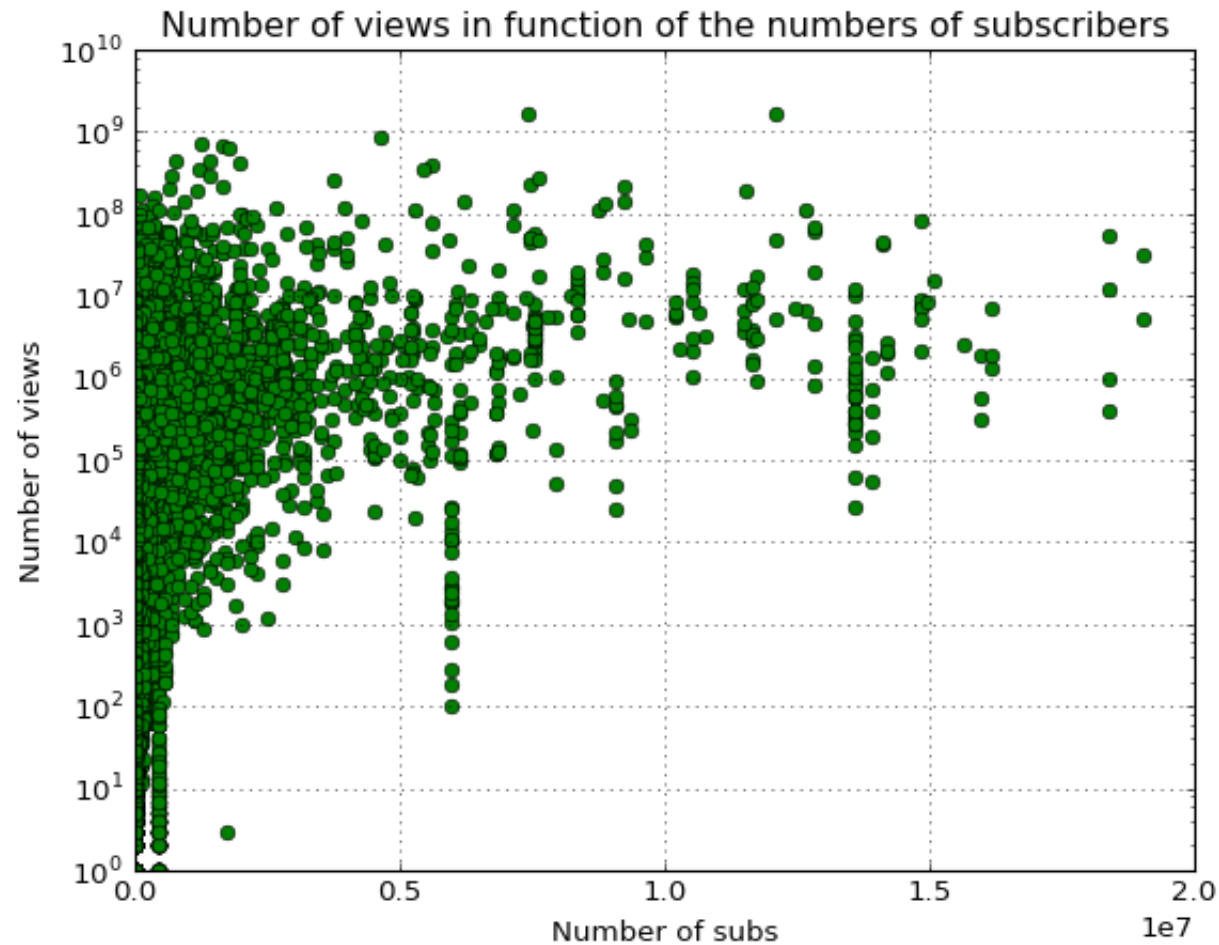 - Title of the video

Data acquisition from Youtube API

- HTTP request
- JSON data format response

Data Exploration

Data Exploration

Neural Network - Thumbnails

- $y = softmax(ReLU(x * W_1 + b_1)W_2 + b_2)$

- 2 layers

- Xavier initializer

- Input: 600-1000 values

- Output: 7 labels

    - [0 - 999]
    - [10000 - 99999]
    - [100000 - 999999]
    - [1000000 - 9999999]
    - [10000000 - 99999999]
    - [100000000 - 999999999]
    - [1000000000 - ∞]

Results

- Theme: random
  - Input : 1000 videos
  - Train accuracy : 0.25
  - 100 test vectors
  - Test accuracy: 0.26

- Theme: cars
  - Input: 600 videos
  - Train accuracy: 0.34
  - 50 test vectors
  - Test accuracy: 0.38

Term Frequency-Inverse Document Frequency (TF-IDF)

- $TF \cdot IDF + 1 = TF \cdot \ln\left(\frac{N+1}{DF+1}\right) + 1$

- Sparse matrix

- Example:
  - *dog* *is* once in all titles, small IDF (1)
  - *cat* and *horse* in 3 titles, higher IDF (1.29)
  - *bird* only in 1 title, highest IDF (1.69)

	bird	dog	cat	horse
« cat cat dog bird horse»	1.69	1	2.57	1.29
« dog »	0	1	0	0
« dog cat horse »	0	1	1.29	1.29

Labels : number of views

- 8 labels for different range of views

- Chosen to have the same number of video per label

- Example: 8049 videos on theme "animals"
  - 0 - 1'004
  - 1'004-2'862
  - 2'862-6'905
  - 6'905-15'632
  - 15'632-36'779
  - 36'779-89'387
  - 89'387-325'975
  - More than 325'975

Neural Network

- $y = softmax(ReLu(ReLu(xW_1 + b_1)W_2 + b_2)W_3 + b_3)$

- 3 layers

- Xavier initializer

- Input: 6000 values

- 1st hidden layer: 1000 neurons

- 2nd hidden layer: 100 neurons

- Output: 8 labels

Results

- Theme: animals
  - Train accuracy : 0.6
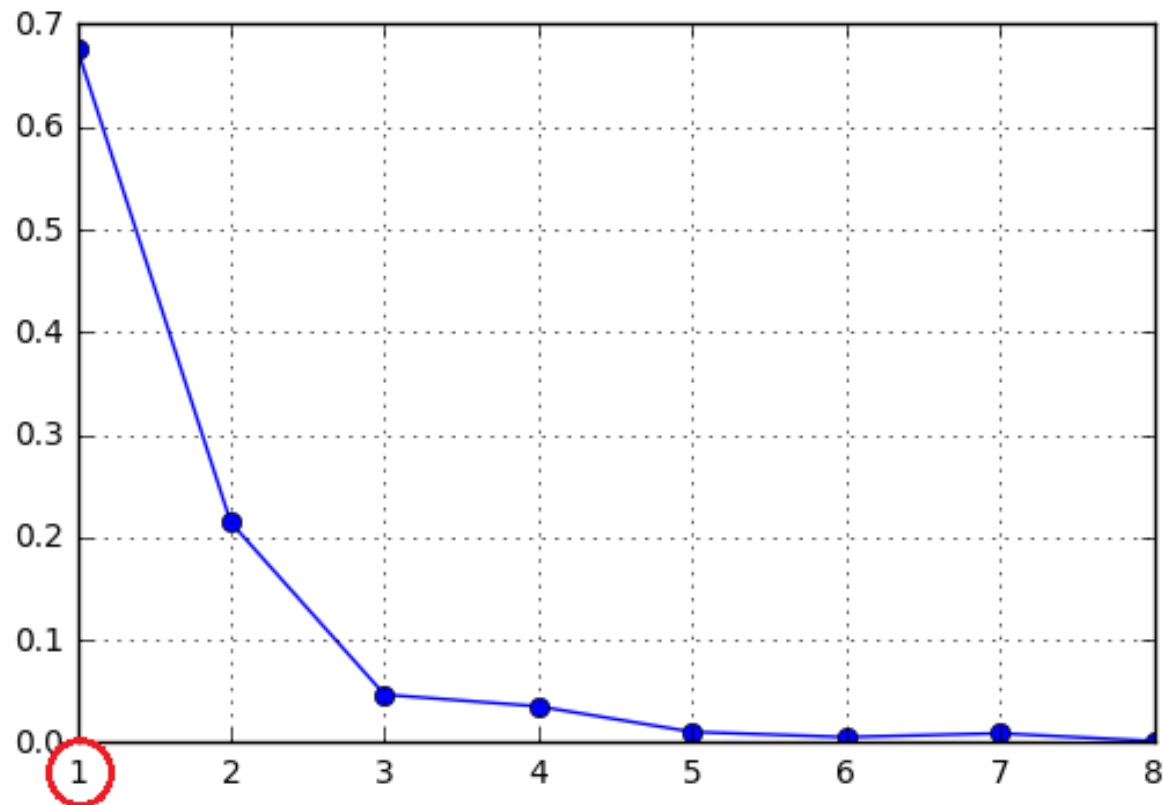 - Test accuracy : 0.3

- Random videos
  - Train accuracy: 0.8
  - Test accuracy: 0.5
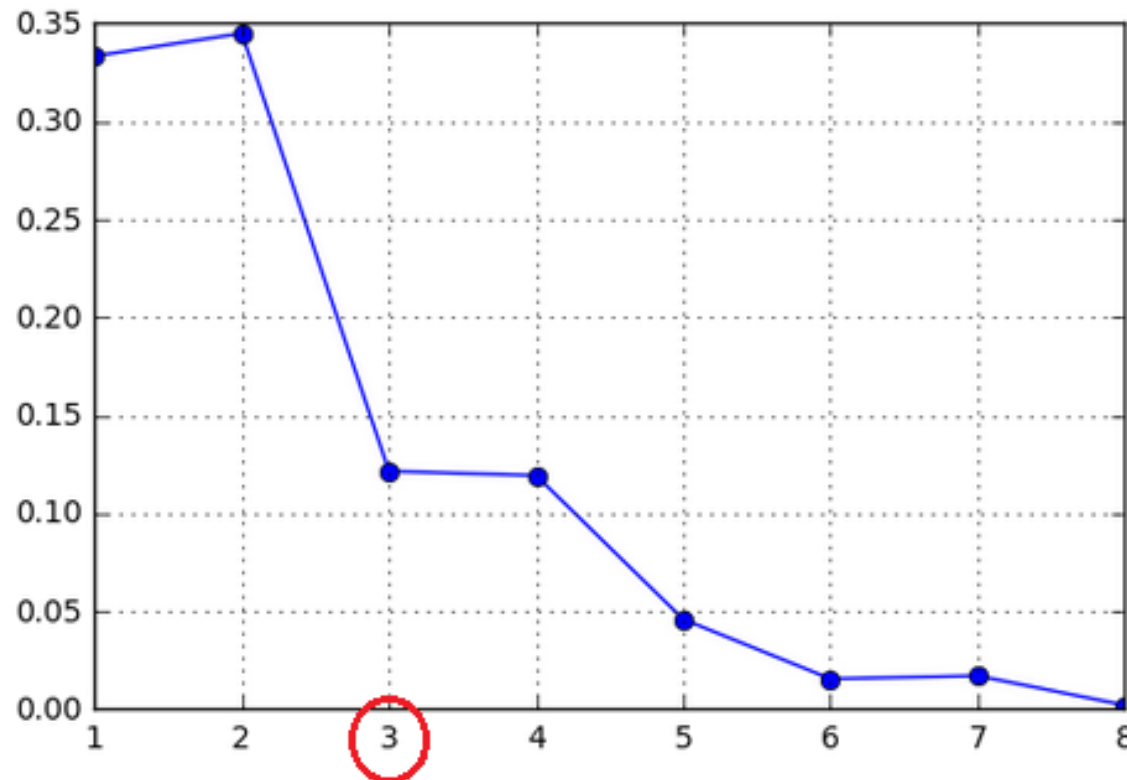
"Un singe en hiver (1962)"

- Subscribers: 2'345   Views: 540

- Label: 0 - 1'004

"Dissection d'un poisson"

- Title: Subscribers: 0   views:3342

- Predicted label: 1005-2862

- Correct Label: 2863-6905

"Les Zinzins de l'espace - Une vie de chien"

- Suscribers: 4'165   Views: 225'648

- Label: 89'388-325'975