



# Memelysis

Analytics of memes

Michał Dyczko  
Jakub Kała  
Maciej Pawłowski  
Krzysztof Spaliński

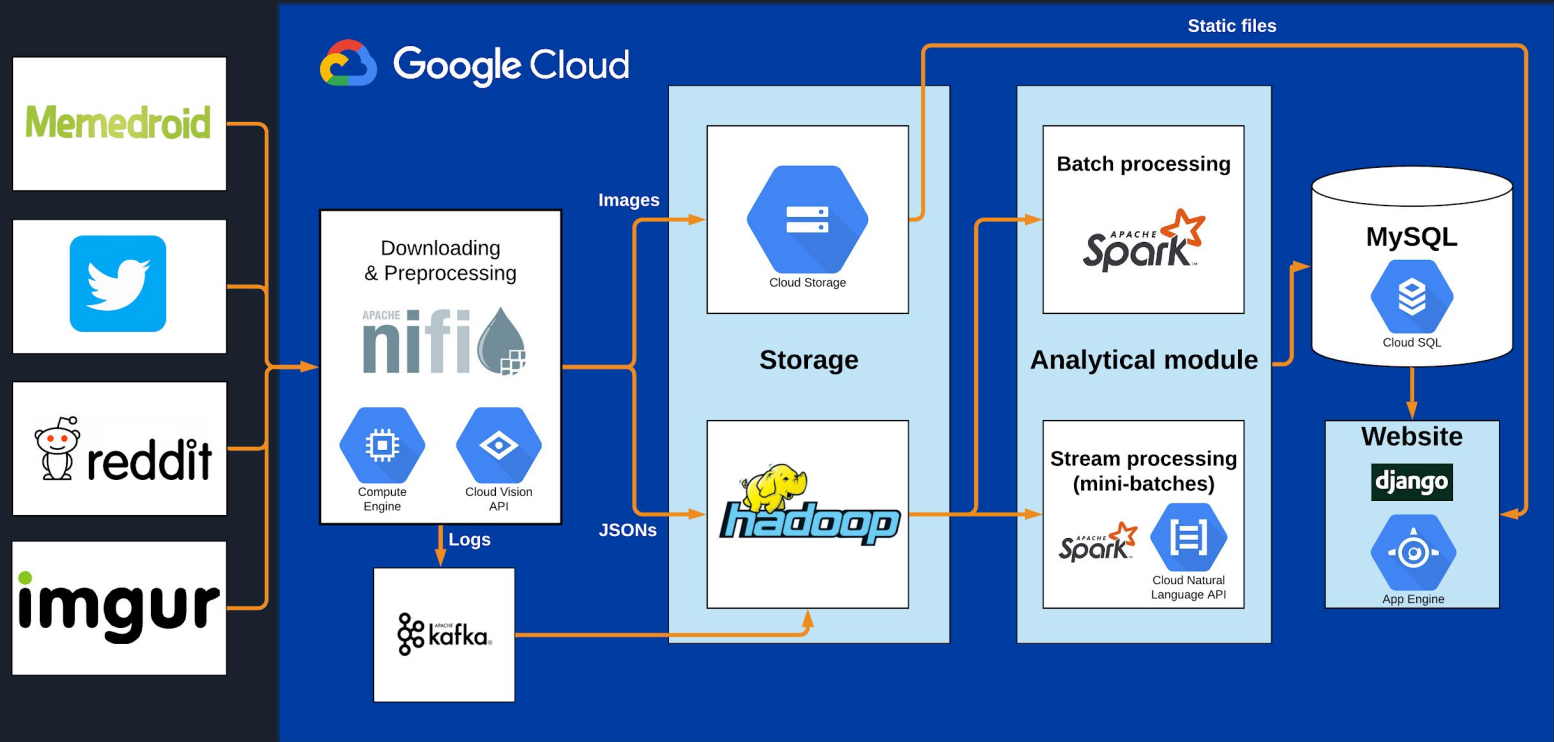


# Project description

The goal of this project is to prepare web-based dashboard consisting of:

1. Live stream of memes from various sources:
  - a. Memes are automatically classified by category;
  - b. Each meme contains information about it's score against other memes from that source.
2. Analytics of meme categories and formats popularity.

# Data architecture





# Data processing

## *Batch processing:*

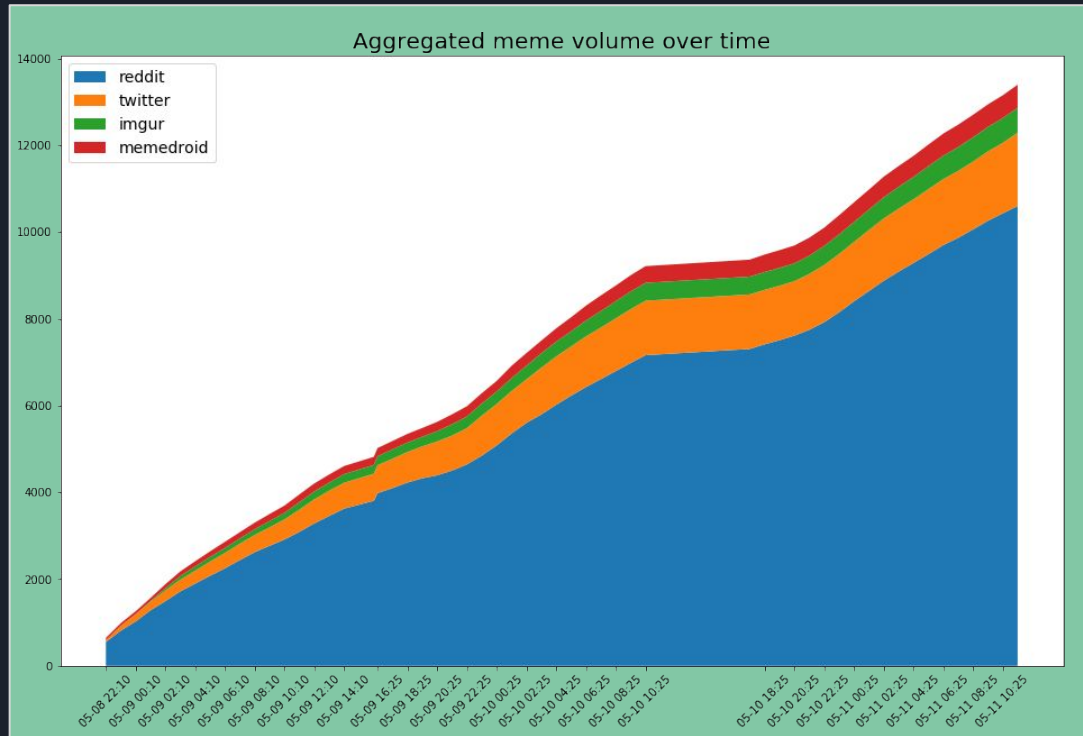
1. Cluster analysis model re-estimation (every 24 hours)
2. Re-estimation of virality factor fitting model
3. Trends analysis

## *Stream processing:*

1. Extracting text from images (OCR)
2. Filtering explicit content
3. Category assignment for each incoming meme
4. **Virality factor** computation



# Data collecting statistics



# Analytical module

Optical Character  
Recognition



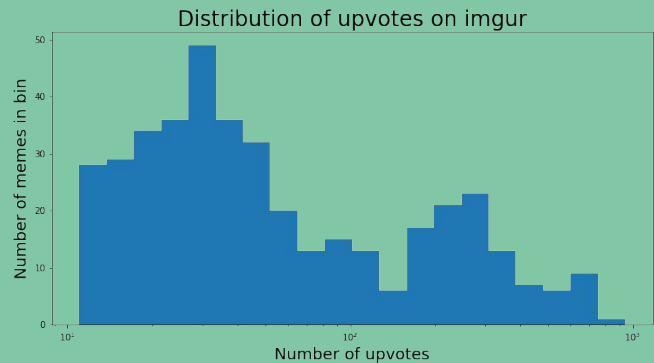
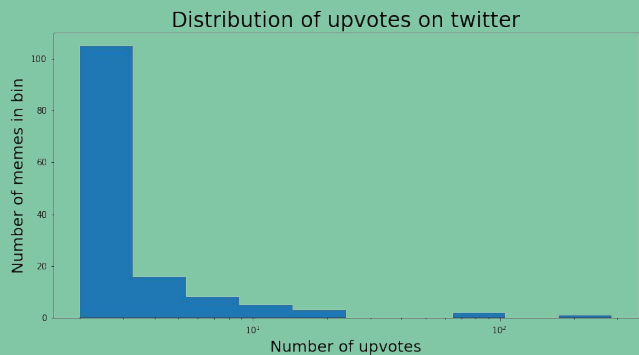
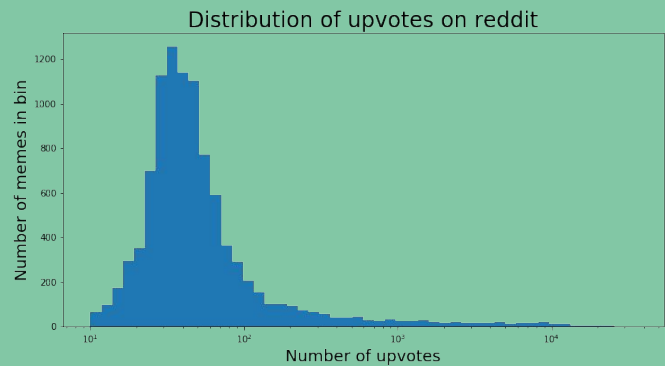
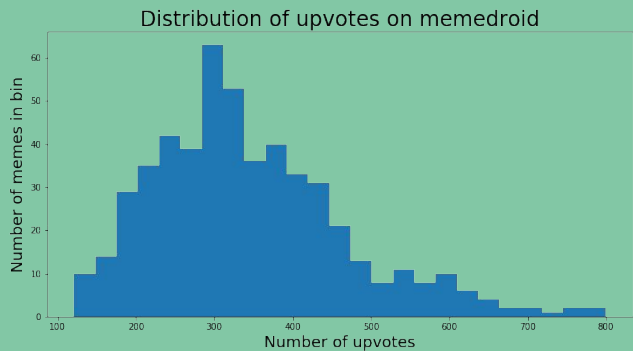
Clustering



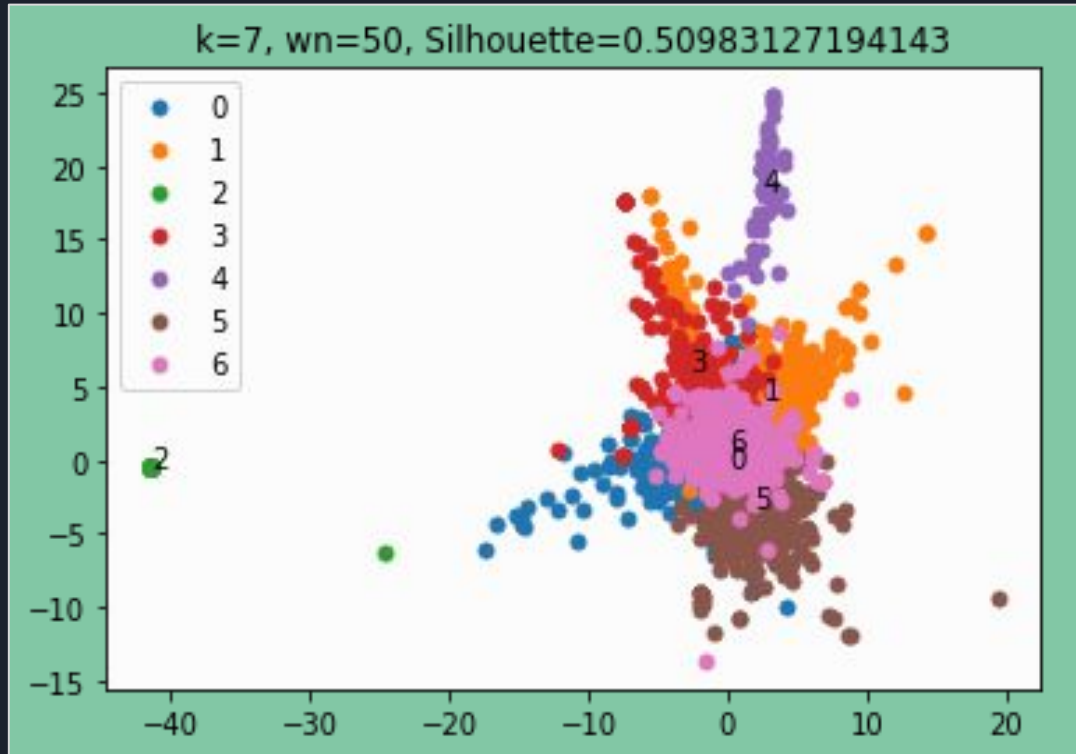
Virality factor analysis



# Upvote distribution on sources



# Sample clustering



id	name
0	funny
1	rpg
2	-
3	-
4	minimumrequirements
5	cyberpunk2077
6	minecraft





# Tools

The Django logo, featuring the word "django" in white lowercase letters on a dark green rectangular background.The Apache Spark logo, with "APACHE" in small grey letters above "Spark" in large grey letters, and a grey star icon to the right.

## Scraping:

- ❖ Python libs: BeautifulSoup 4, requests, Django 3;
- ❖ API's: Twitter API, Reddit API.

## Cloud - GCP:

- ❖ Compute Engine running Nifi for scraping and image downloading;
- ❖ Dataproc: Hadoop and Spark;
- ❖ Google Storage for image hosting;
- ❖ App Engine running Django app.

The Apache NiFi logo, with "APACHE" in small grey letters above "nifi" in large grey letters, and a blue water drop icon with a grid pattern to the right.

## Analytics:

- ❖ OCR - Google Cloud Vision API.

The Apache Hadoop logo, featuring a yellow elephant icon to the left of the word "hadoop" in large blue letters, with "APACHE" in small red letters above it.

# Product

## Sample object

The potato I peeled:



The potato my mother peeled:



Meme from  
reddit posted at  
May 8, 2020, 4:08  
p.m.

Click for more

