Politechnika Wrocławska

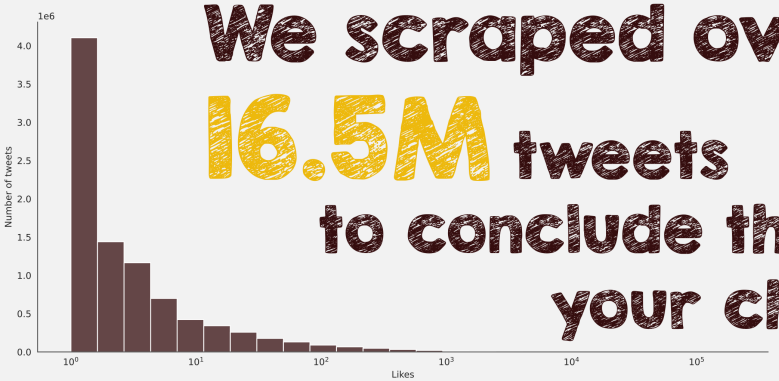
Karol Gawron | Michał Pogoda | Norbert Ropiak | Michał Swędrowski

PROJECT PSALM

Predict SociAL Media

How do you measure popularity?

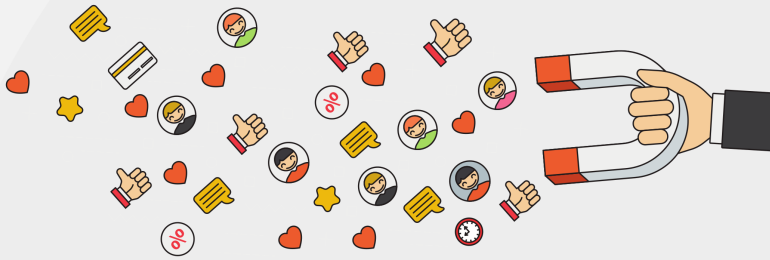
AFAIR is metric designed with a view to represent popularity of given tweet. It consists of two components - AFR and AIR. AFR stand for Average Followers Ratio and it represent all tweet interactions taking into account followers of a user. AIR means Average Internal Ratio which depicts deviation of actions from average number of curren user actions. AFAIR is sum of normalized AFR and AIR.



We scraped over 16.5M tweets to conclude that your choice of words is

Feel lonely?

Not-so-fun fact: Only 52% of tweets we scraped had one or more likes



Church | LGBT | Ecology | Medicine | School | Music | Sport (Legia) | Credit | Beauty | "I know" | No/sorry

However, we found different **topics** in **popular & unpopular** tweets

Popular topics tend to be related to worldviews - touching the most sensitive aspects of private life (Church, LGBT). Additionally, it is worth to notice it is also connected to potential fields of legislation changes (Ecology, School, LGBT). Pop culture themes are also very popular (Music, Sport). On the other hand, non-popular topics look like they could be the subject of a private conversation. It is no surprise that posts with only one recipient do not generate much action. Influencers are not used to getting into personal disputes. In summary, the most popular topics often evoke very strong emotions in the audience. Another important factor is that the topic can be understood by a wide spectrum of users.

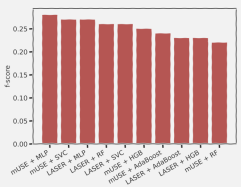
for tweet popularity

How we wanted to predict?

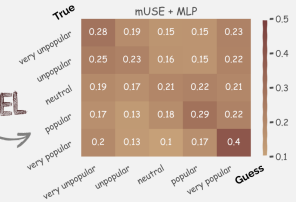
We wanted to use SOTA models such as well-known BERT or LASER. Due to language knowledge contained in the base model we thought, we will be able to analyze dependencies between parts of a text. Thanks to the fine-tuning of the models, we were trying to specialize our model in terms of popularity classification. Then, owning a final model, we used the Captum library, which allowed us to examine the attribution of individual words on the prediction score. This allows us to tell which words affect the popularity of a given text.

What we got?

The results are not impressive. Regardless of the approach used, we were not able to obtain a satisfactory model performance. Prediction precision achieved, on average, less than 30% on five classes of popularity shows that the model is not much better than the random approach. We can see an interesting trend of increasing prediction confidence for the class representing very popular tweets. The results show that the choice of words might not have any significant influence on the popularity of a post.



BEST MODEL



Still want to check if your tweet will go viral?

Go to the site and see why your post will be popular*

*At least why our model thinks that way - it's not that good at predicting...