



Edyta Rogula



Marta Maślankowska

MEMES

Monotonous Evaluation of Memes from Internet Scraping



Witold Kościukiewicz



Łukasz Sus

MEMES

Meme - an image, video, piece of text, etc., typically humorous in nature, that is copied, modified and spread rapidly by Internet users.

WHAT ABOUT THEM?

- Are offensive or non-offensive memes more popular?
- Do people like racist or non-racist memes more?

OUR MAIN TASK

Classify memes in three binary categories: offensiveness, inappropriateness and racism.



DATA GATHERING

We collected English memes from ImgFlip. Memes were annotated using our own mobile app. We solved annotating disagreement solution with test questions.

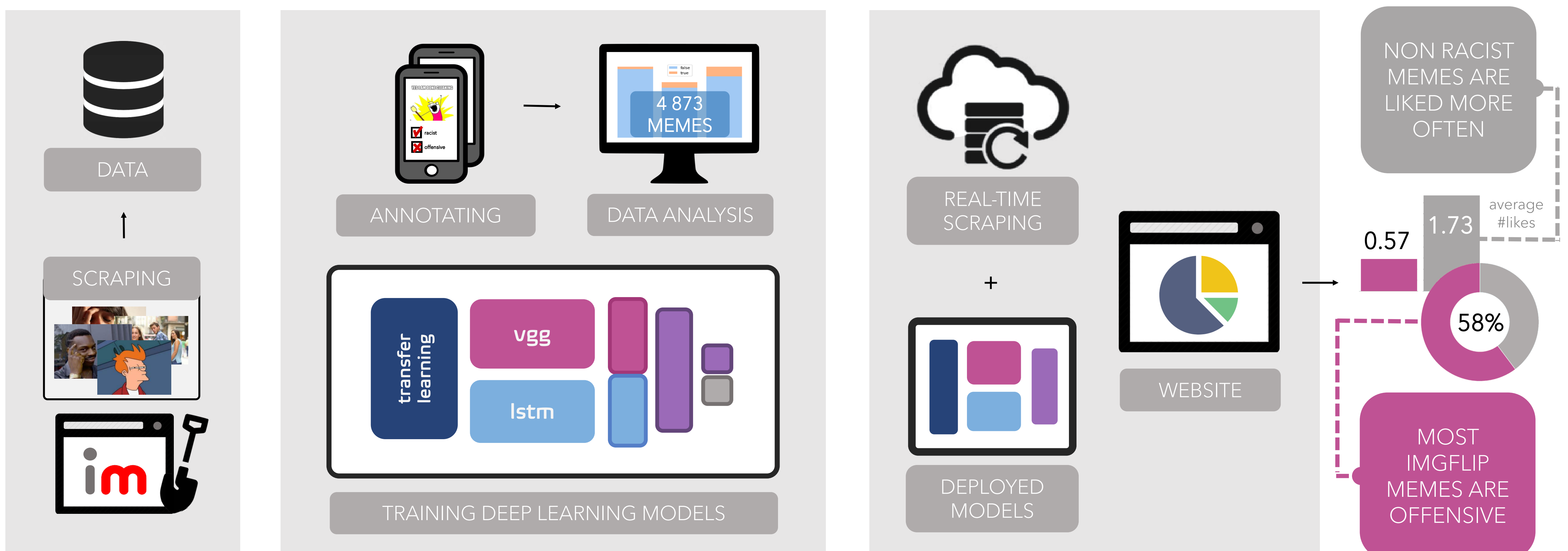
DEEP LEARNING STUFF

Model: pre-trained LSTM (texts) & pre-trained VGG (images)

Result: meme embedding representation

TRANSFER LEARNING

We used external texts datasets of hate speech and toxic comments (more than 150 000 tweets) to pre-train the language part of our model. VGG model was trained on ImageNet.

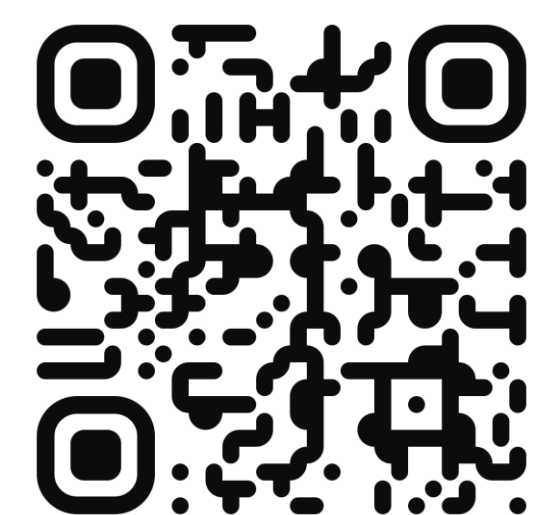


ENCOUNTERED PROBLEMS

- ✗ slang language and abbreviations
⇒ fasttext embeddings covering unknown words
- ✗ data imbalance (only 5.2% of racist memes)
⇒ oversampling
- ✗ annotation: memes' ambiguity
⇒ additional test expressions unifying understanding

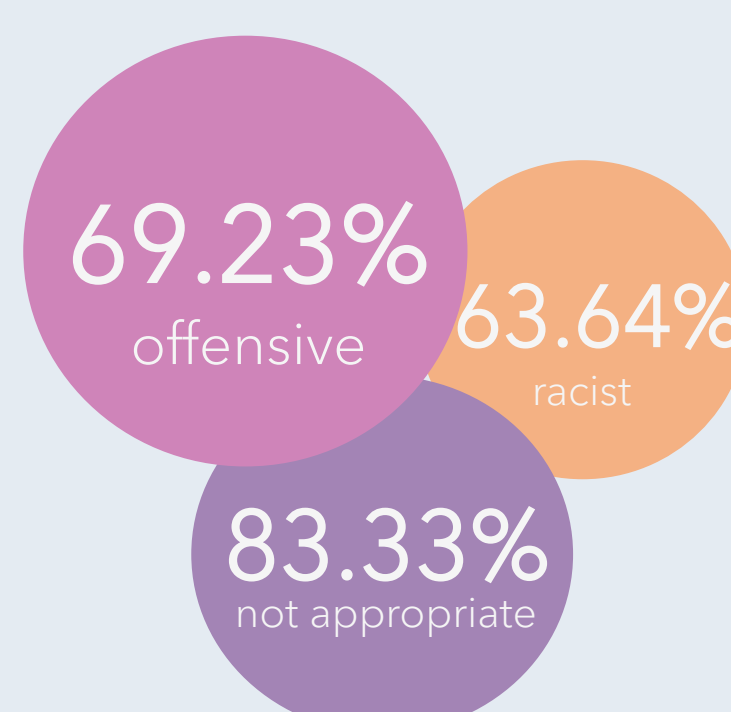
WEBSITE

We prepared a website with continuous analysis of periodically scraped memes. Our deployed model categorizes them and obtained results are presented in human-friendly visualizations.



PROEJCT'S RESULTS METRICS

- precision for offensive: how many memes selected as offensive are really offensive?
- recall for inappropriate and racist: how many relevant memes are selected?



TAKEAWAY THOUGHTS

- we created the first **automated real-time analysis** of offensiveness of Internet memes
- data annotation is very difficult but crucial for satisfying results
- everyone can create a meme and share it; it makes them as diverse as we are