

# PEAP

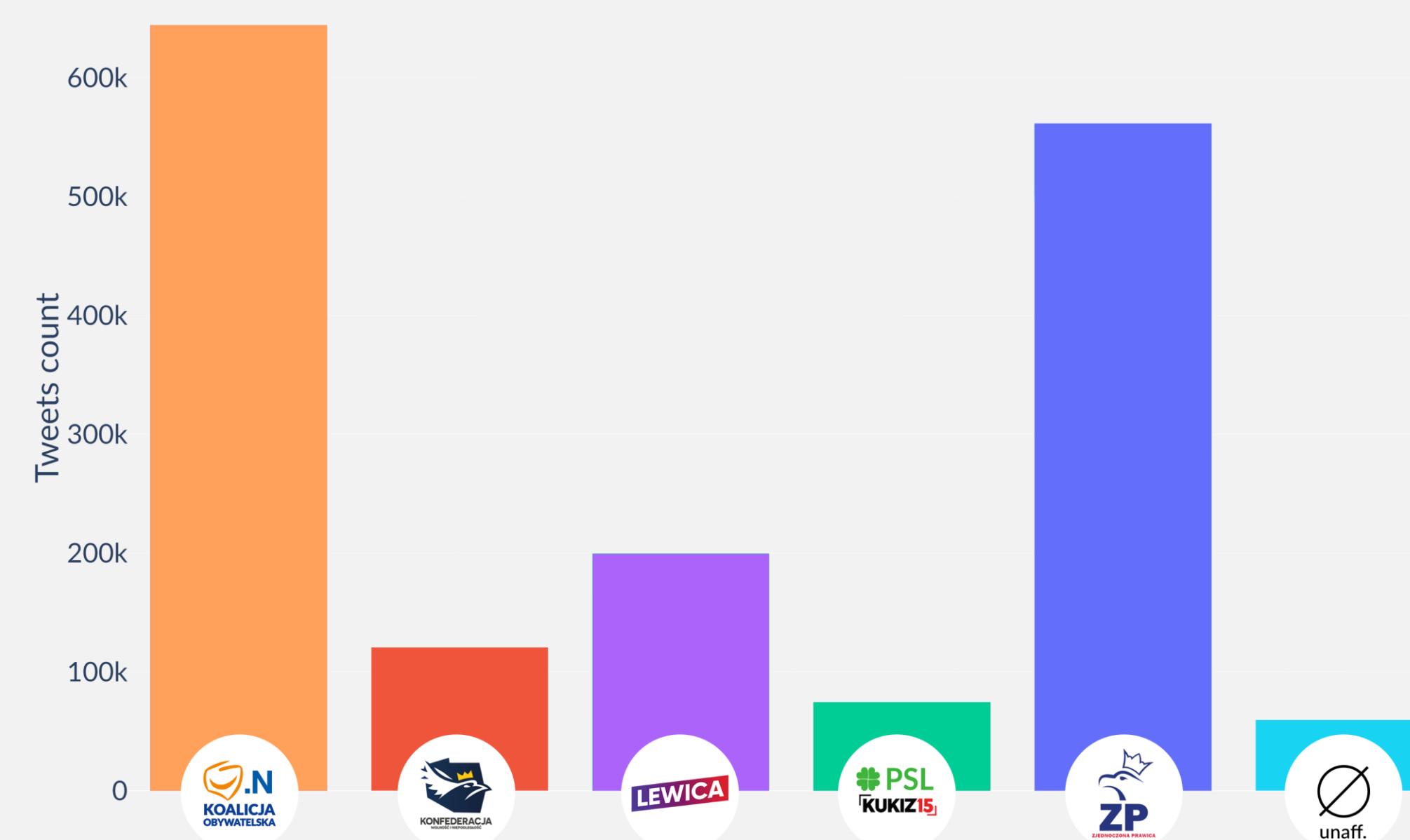
Politicians Exploratory Analysis Platform



## Idea

Latest polls and election results show that Poles' view on politics is highly polarized. We think that politicians' tweets strongly reflect their opinions on various issues. With this in mind, we gathered over **1.6M** tweets of **537** Polish politicians and analyzed them in terms of topic and sentiment. We used the results to create an interactive platform just for you to browse it.

## Data



We manually created a list of Polish politicians' Twitter accounts and identified **581** accounts from **19** parties, forming **5** coalitions. We also included unaffiliated users in our dataset. We gathered all tweets since account creation.

All data was collected in November 2020.

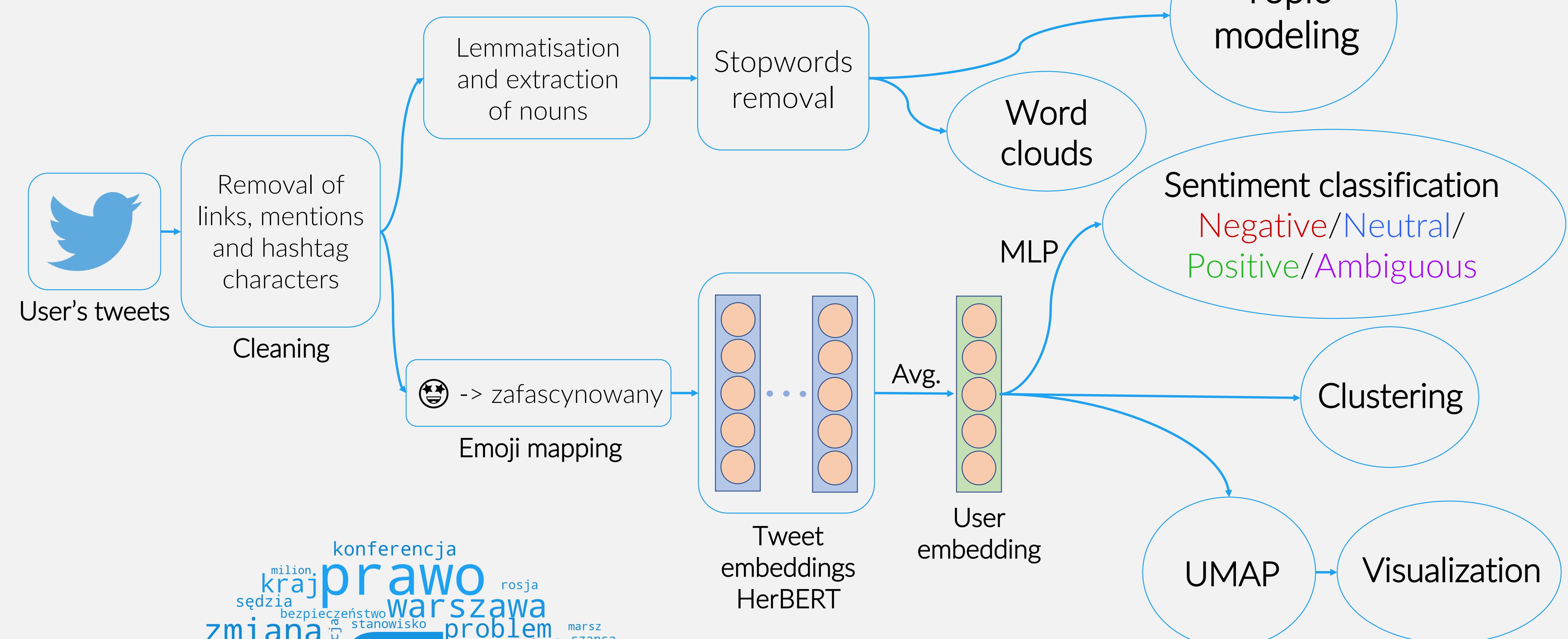
## Topic modeling

We used Latent Dirichlet Allocation (LDA) to model topics of tweets. We trained model on all our tweets and identified **10** topics. On our platform we provide words distribution for each topic to better explain their meaning.



We provide similar analysis for every user, party and coalition. Check them out on our platform and dive into analysis!

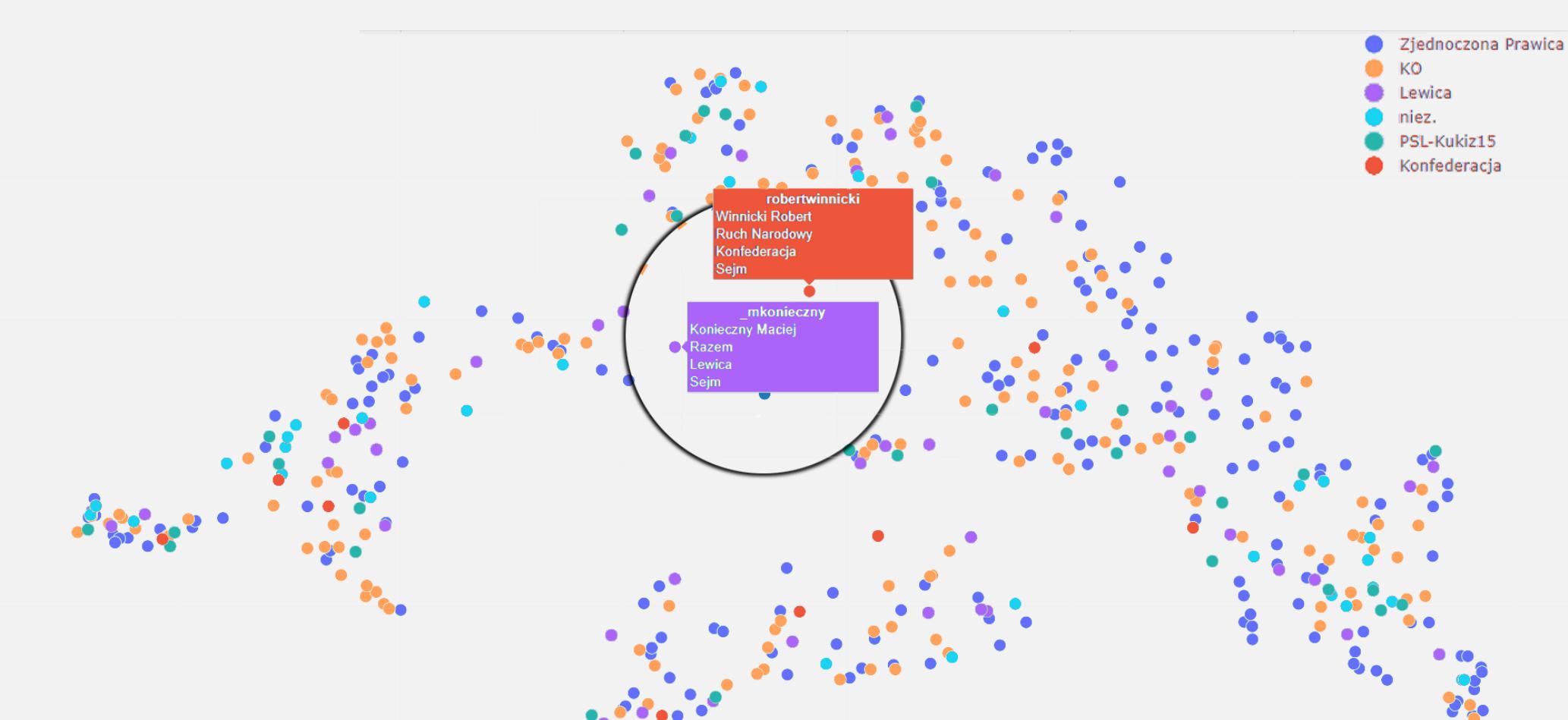
## Pipeline



Most frequent words

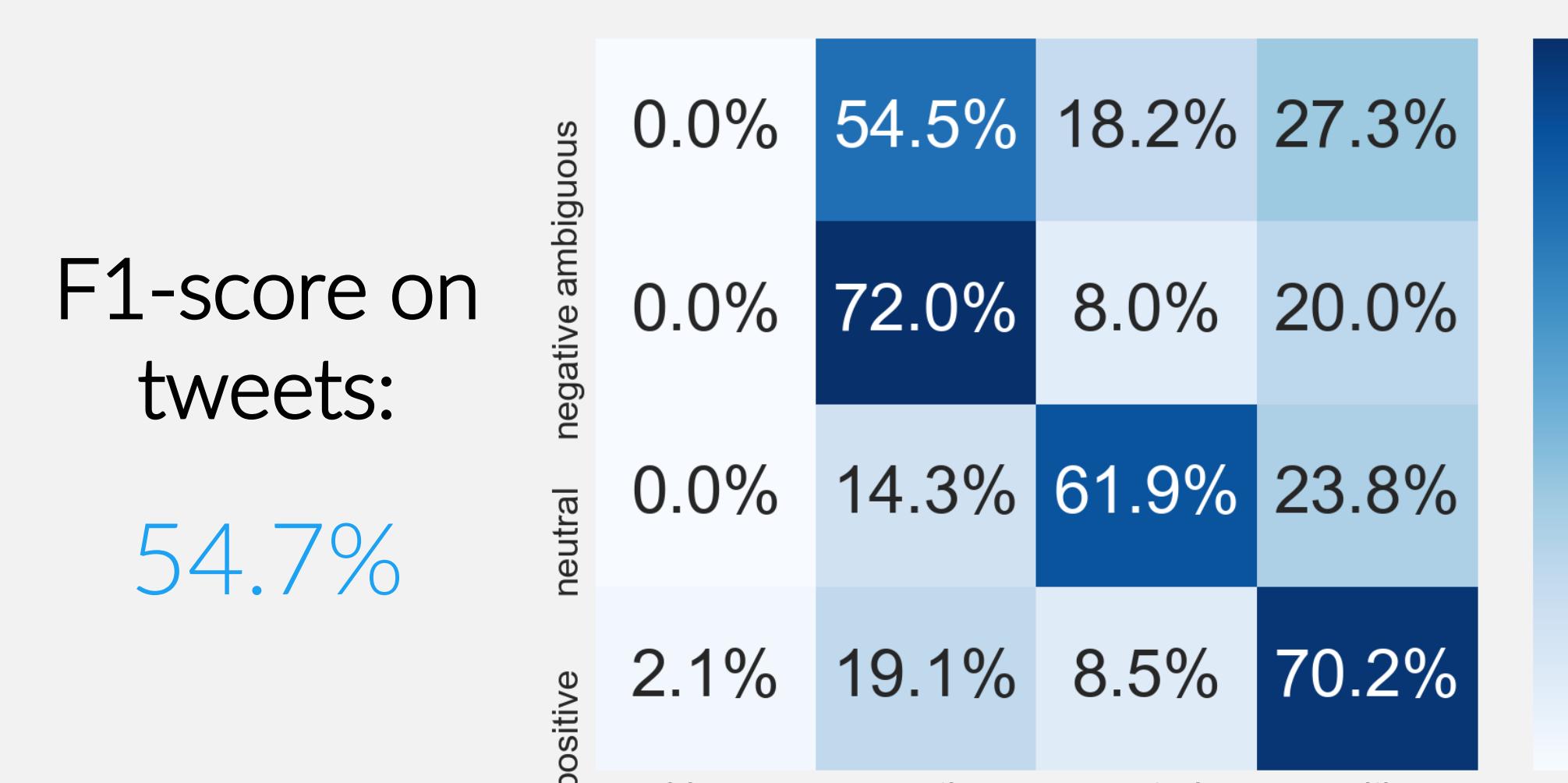
## User embeddings

We used a fine-tuned HerBERT model to embed each tweet in a vector space. We averaged tweet embeddings for each user to get their embedding. We then used several clustering algorithms to group users and UMAP to allow us to present the results in 2 and 3 dimensions.



## Sentiment

We annotated **1038** randomly chosen tweets to one of four classes: negative, neutral, positive or ambiguous. We achieved Cohen's Kappa - **0.632**. Our dataset consists of PolEmo dataset, annotated tweets and sentiment examples gathered from PLWordNet. We trained a neural network for sentiment classification.

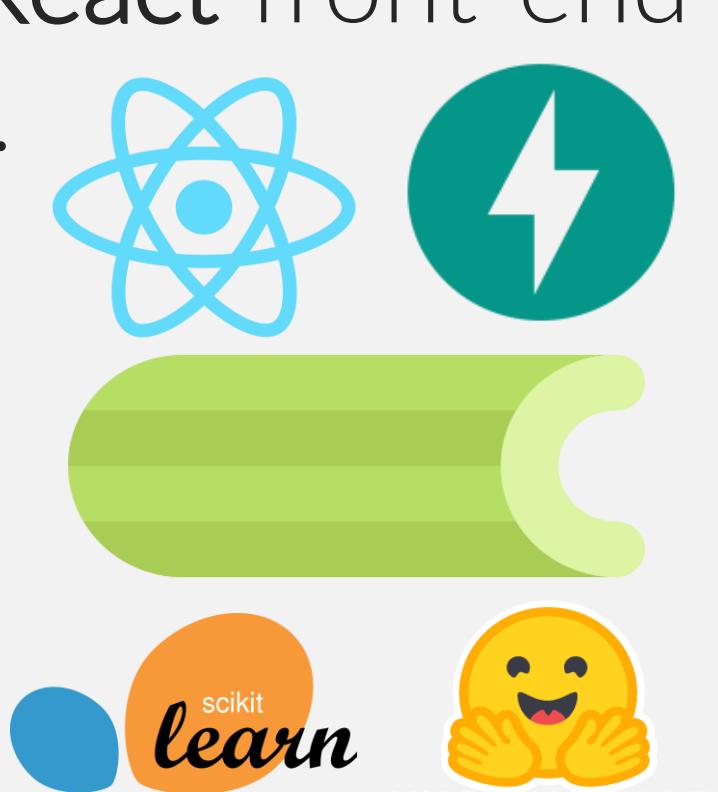


Contact: embedd.ml@gmail.com

## The Platform

Every analysis can be found on our platform. As a bonus, you can analyze recent activity on any account you want!

Our app was created with a React front-end side and a FastAPI backend. The ML part is handled by Scikit-learn as well as Hugging Face. We made use of Celery to run our account processing pipeline.



## Check it out for Yourself!

<https://politicians.embedd.ml>

