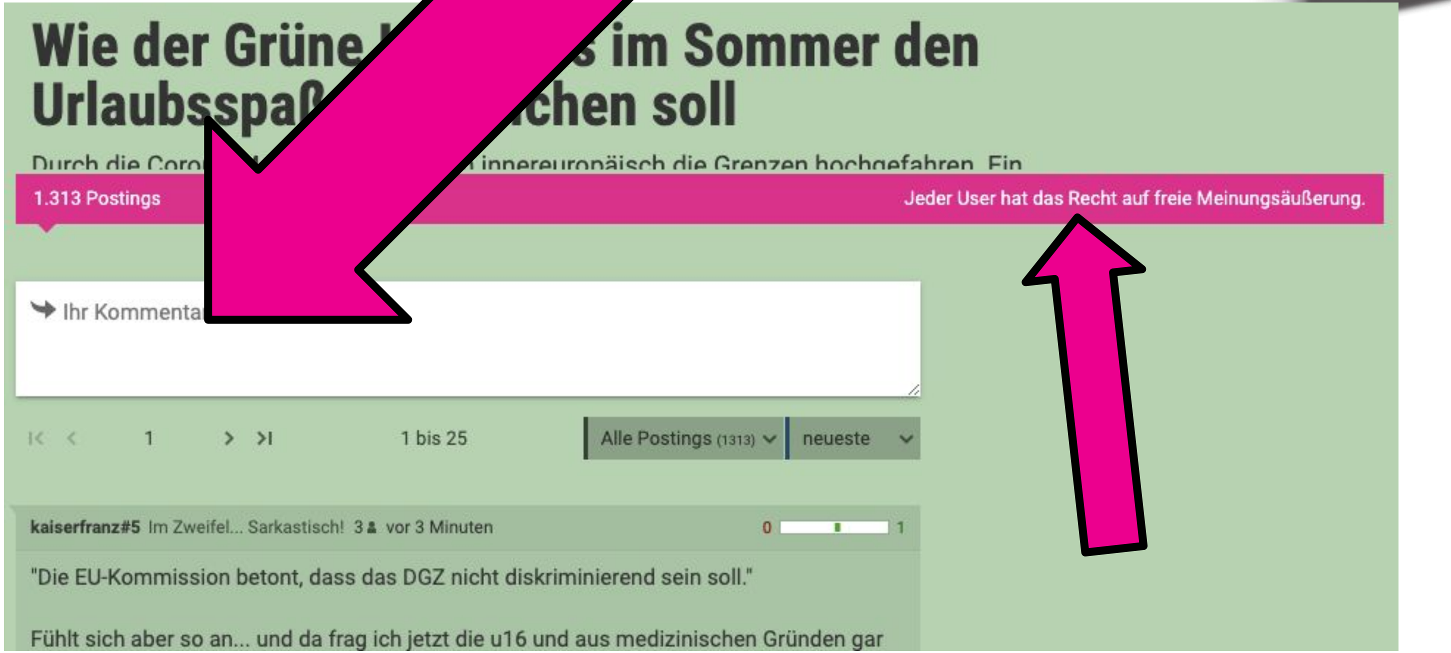
One Million Posts

Der Standard

chrsteck, dominikmn, lima-tango

Der Standard



Wie der Grüne … s im Sommer den Urlaubsspaß … chen soll

Durch die Coro… innereuropäisch die Grenzen hochgefahren. Ein

1.313 Postings

Jeder User hat das Recht auf freie Meinungsäußerung.

Ihr Kommenta…

kaiserfranz#5 Im Zweifel... Sarkastisch! 3 vor 3 Minuten 0 1

"Die EU-Kommission betont, dass das DGZ nicht diskriminierend sein soll."

Fühlt sich aber so an... und da frag ich jetzt die u16 und aus medizinischen Gründen gar

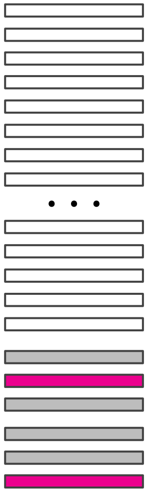
Task

Help the newspaper:

- Detect potentially problematic user posts
- Predict if a post …
  - … is discriminating
  - … uses inappropriate language
  - … has a negative sentiment
- Reduce moderation expenses

Dataset

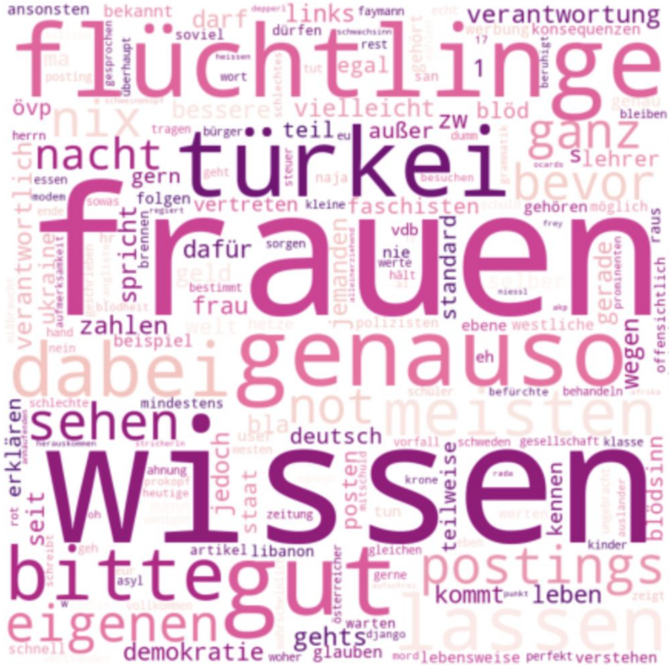
- >1 Million user posts from ‘Der Standard’
- 1000 labeled posts (randomly from 1 year)
- 3+1 labels
    - discriminating
    - inappropriate
    - sentiment negative
    - needs moderation



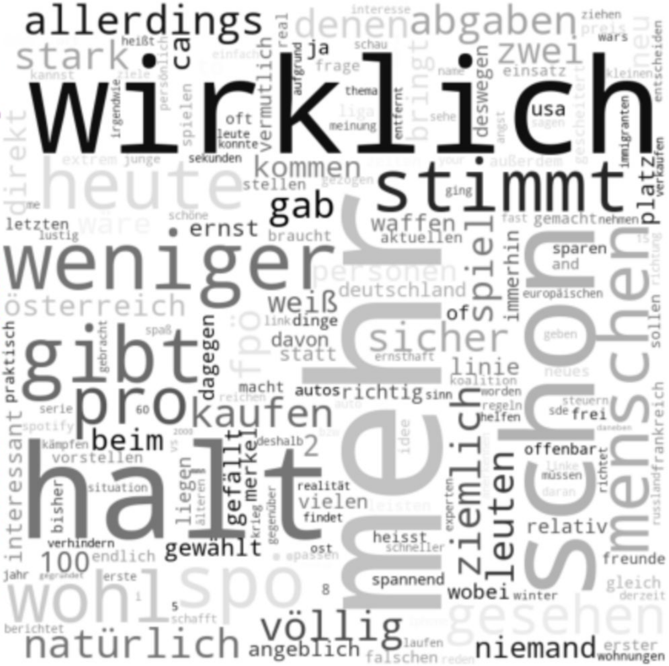
EDA

Exploratory Data Analysis

EDA - Typical words for the labels



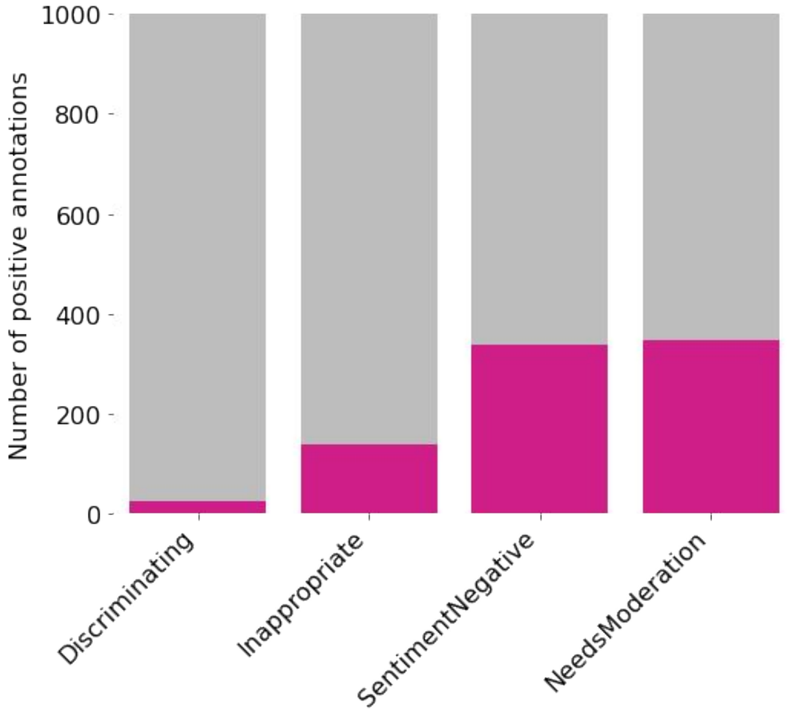
Needs moderation



EDA - Positive annotations per label

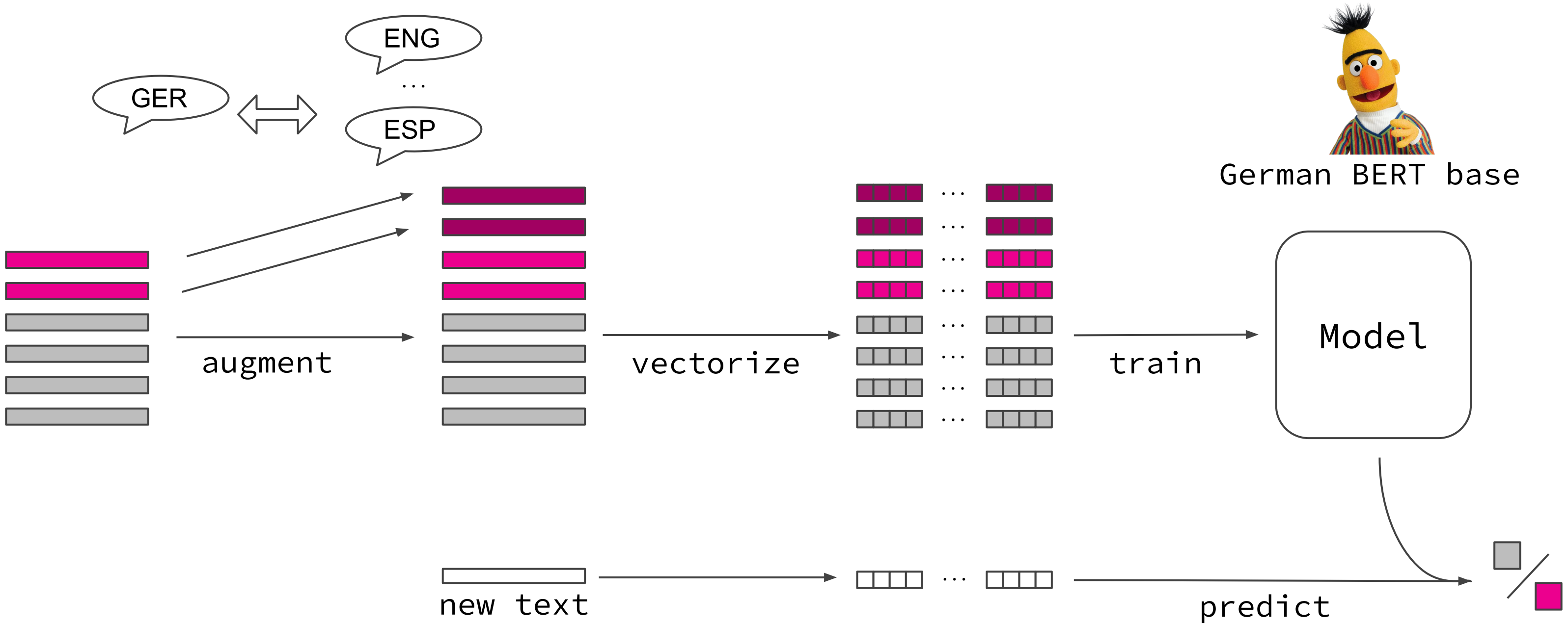
Imbalance in categories:

- Discriminating < 3%
- Inappropriate 14%
- Needs moderation & sentiment negative > 30%



Modeling

Architecture



ENG

…

GER

ESP

German BERT base

augment

vectorize

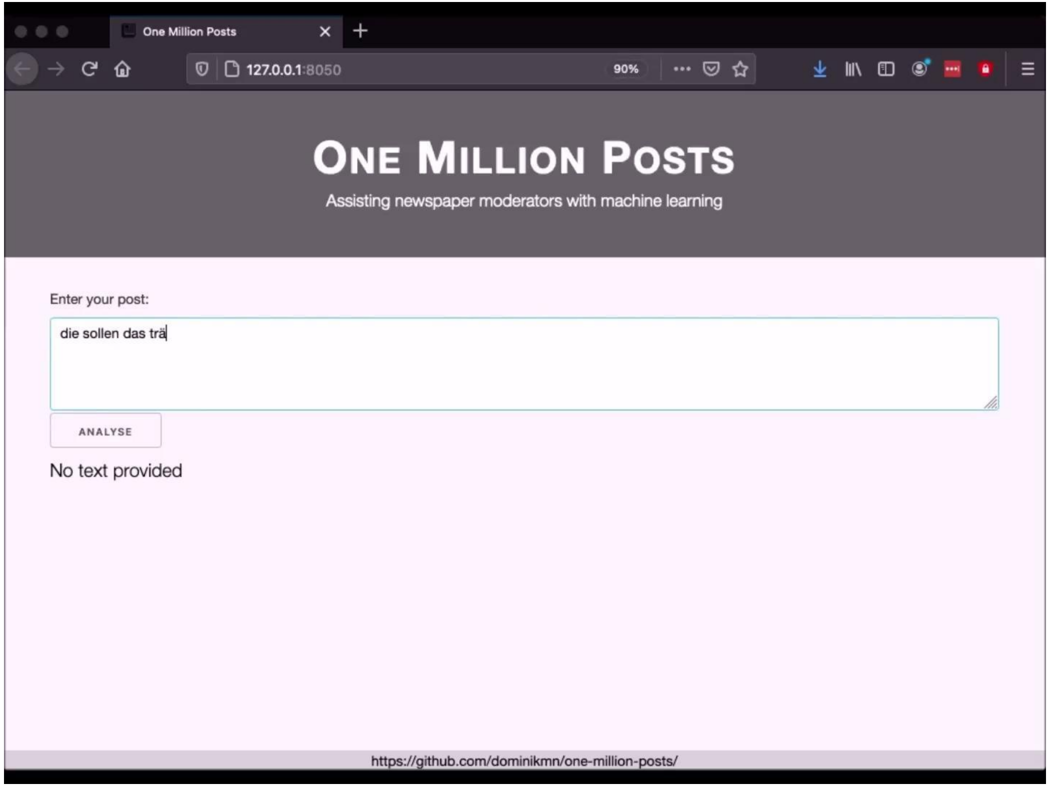
train

Model

new text

predict

Application



ONE MILLION POSTS

Assisting newspaper moderators with machine learning

Enter your post:

die sollen das trä

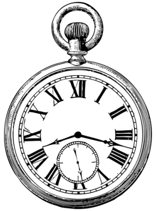
ANALYSE

No text provided

https://github.com/dominikmn/one-million-posts/

Achievement

37% less moderation effort



94% of problematic comments detected

Alternative models between:

- 52% reduced effort / 84% problem detection
- 16% reduced effort / >99% problem detection

Future work & Experience

Future Work

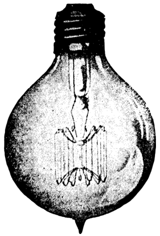
Manual annotation of unlabeled data

Semi-supervised learning on 1 million posts

Train BERT with **MLM** (Masked Language Model task)

Tune model towards **fully automated** moderation

Experience



Consequent product focus -> Clear target for your project

Team workflows -> Kanban board, git conventions, MLflow

Data centric approach -> A model is only as good as its data

Unexpected scores -> Can expose errors in your pipeline chrsteck, dominikmn, lima-tango

https://github.com/dominikmn/one-million-posts