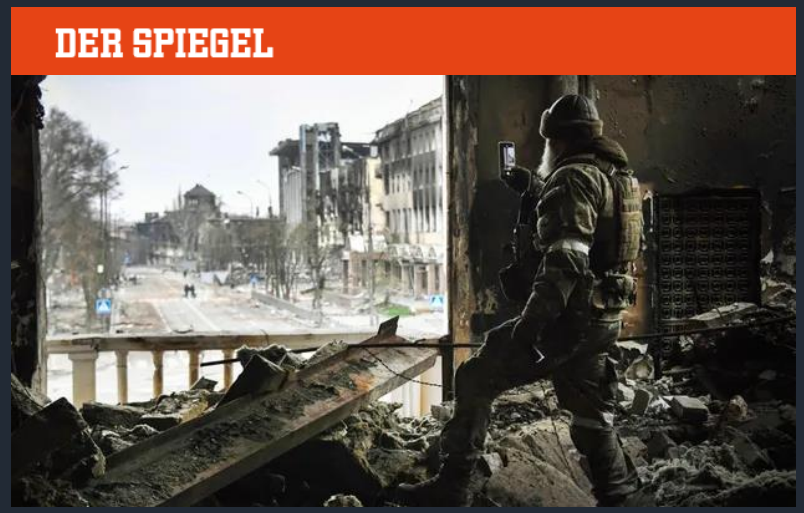
Predicting Newspaper Publishers based on Article Images and Texts



Fabian Paul, 10/08/2022

Business Problem

Die Militäroperation gegen den „Islamischen Dschihad" im Gazastreifen wird in Israel als Erfolg gewertet. Die Gruppe blieb in ihrem Kampf allein – die Hamas beteiligte sich nicht an der jüngsten Eskalation.

- Insights on images and texts crucial for:
  - … journalists to **deliberately choose motives**
  - … readers to **reflect on choice of motives** of different publishers

Nach inszenierter Habeck-Entführung

Faeser kündigt harte Konsequenzen bei Bedrohung von Politikern an

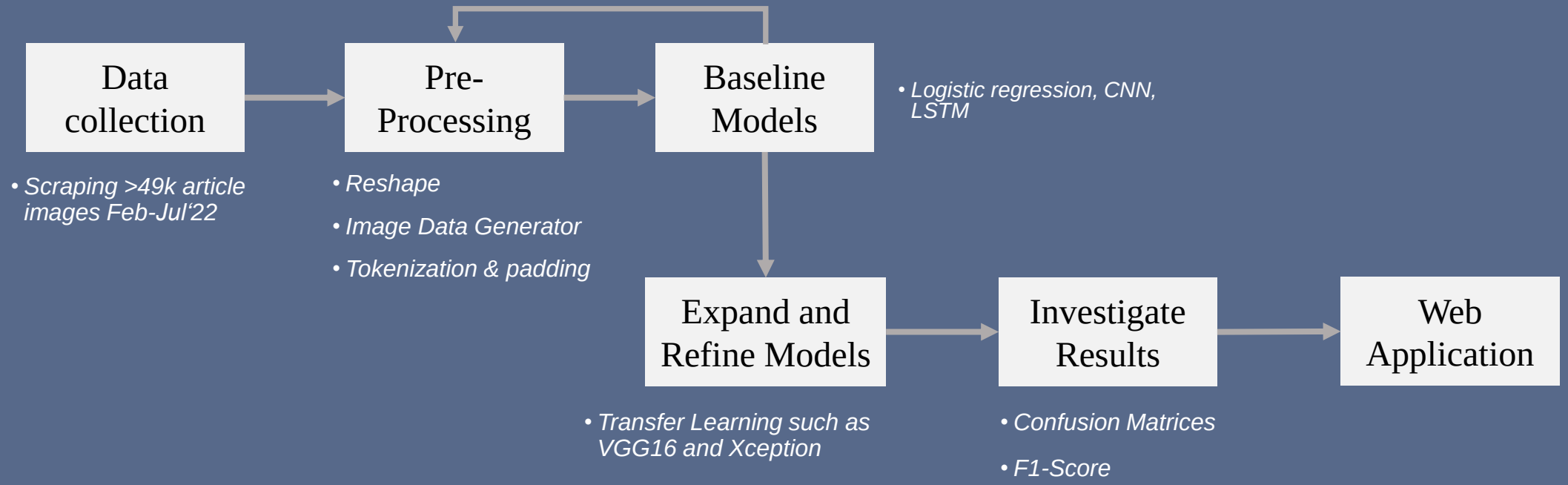
Rechtsextreme werben mit einer gestellten Entführung von Wirtschaftsminister Habeck für eine Demo – und berufen sich auf die Kunstfreiheit. Nun hat Innenministerin Faeser klargestellt, wo das Recht auf Protest endet.

2 Min

Objective

Can we systemtically and reliably **differentiate images and teaser texts** between different publishers?

Methodology

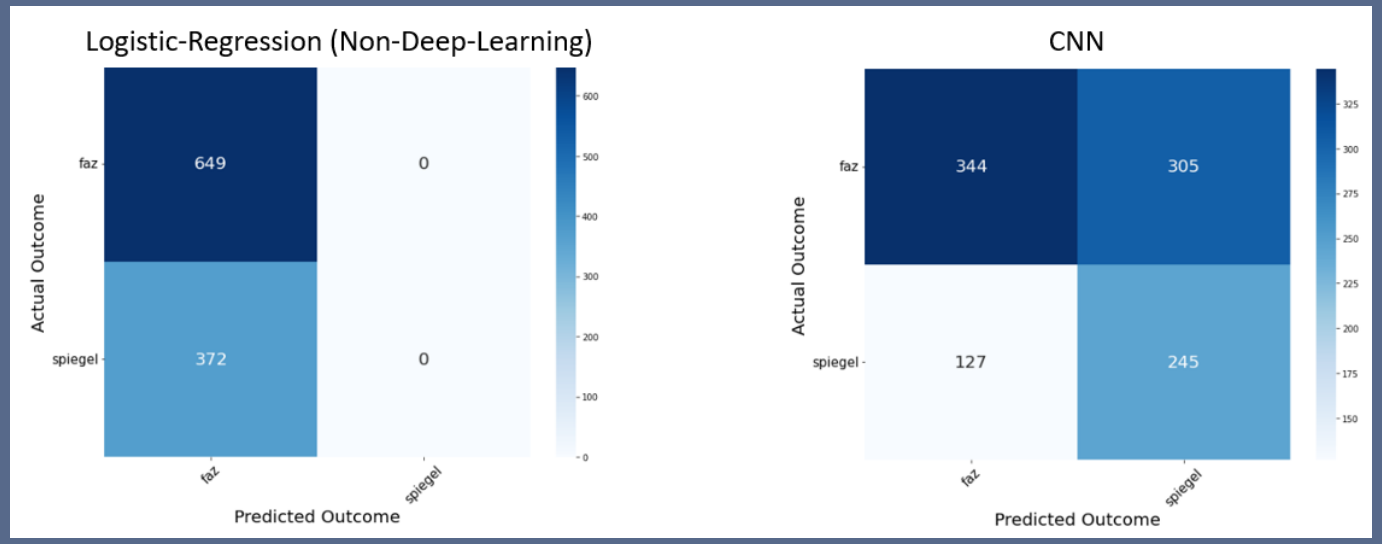


Results – Multiclass Image Prediction

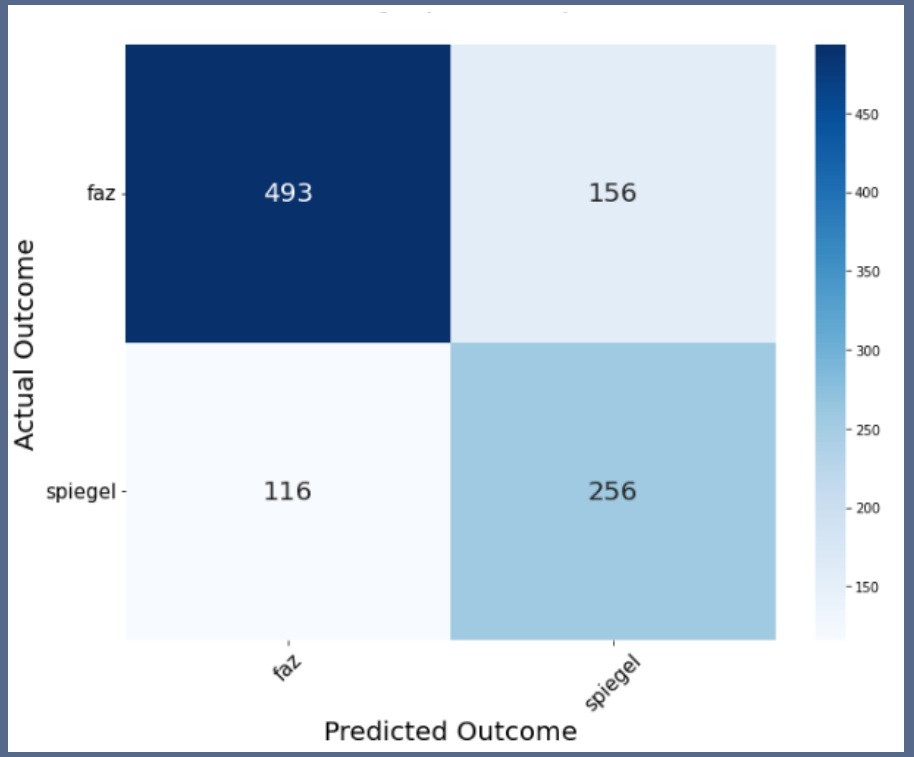
Model	Accuracy	Recall	Precision	F1 Score
CNN Base	0.39	0.34	0.36	0.32
Mobilnetv2 Base	0.40	0.40	0.40	0.39
Mobilnetv2 Trainable	0.34	0.25	0.09	0.13
Xception Base	0.41	0.37	0.43	0.35
Xception Trainable	0.42	0.40	0.40	0.40
VGG16	0.44	0.41	0.41	0.41

Results – Binary Image Prediction

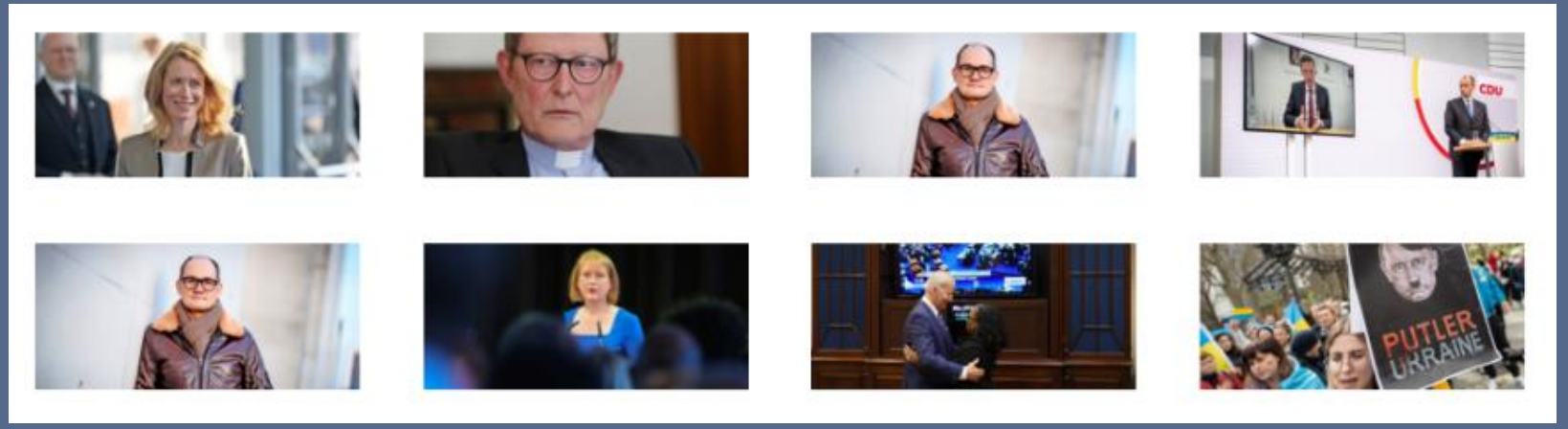
Baseline Models



VGG16 Transfer Learning Model

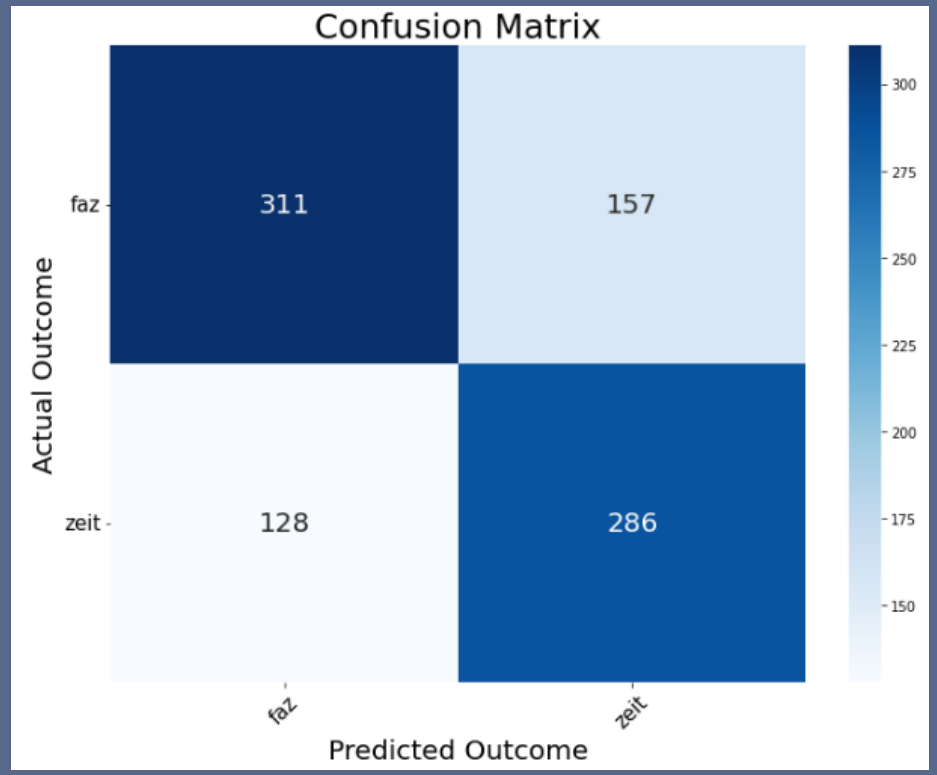


Results – Binary Image Prediction



Results – Binary Text Prediction

Baseline LSTM




```
accuracy : 0.6768707482993197
recall : 0.6770132200722965
precision : 0.6776755852842808
f1 : 0.6766109309287673
```

Conclusion

Results

- Medium performance: Deep Learning Models for multiclass image classification
- Good performance: Deep Learning Models for Binary Image and Binary Text classification

Link Web Application

Limitations

- **Limited size of datasets**: Approx. 4 tsd. per publisher per category
- **Focus category** of dataset: Politics
- No **transfer learning for text** classification
- Combine **image and texts in joint deep learning model**
- Performance increase by **usage of GPUs**

Future Work

1. →Increase size of image and texts datasets
2. →Expand analysis on more than one category
3. →Include pre-trained German NLP deep learning models
4. →Develop joint model for image and texts
5. →Increase performance by using GPUs

Thank you for your attention!