

HumAID: Human-Annotated Disaster Incidents Data from Twitter with Deep Learning Benchmarks

Firoj Alam, Umair Qazi, Muhammad Imran, Ferda Ofli
Qatar Computing Research Institute (QCRI), HBKU, Qatar



Highlights

- A new large-scale dataset with ~77K human-labeled tweets
- Sampled from a pool of ~24 million tweets across 19 disaster events that happened between 2016 and 2019
- We propose a data collection and sampling pipeline, important for social media data processing
- We report multiclass classification results using classic and deep learning (fastText and transformer) based models
- Publicly available at https://crisisnlp.qcri.org/humaid_dataset.html

