

NLP NORTH AT WNUT-2020 TASK 2: PRE-TRAINING VERSUS ENSEMBLING FOR DETECTION OF INFORMATIVE COVID-19 ENGLISH TWEETS

IT UNIVERSITY OF COPENHAGEN

Anders Giovanni Møller • Rob Van Der Goot • Barbara Plank
IT University of Copenhagen



Problem: Binary classification of INFORMATIVE and UNINFORMATIVE corona-related tweets provided by the organizers.

Code: https://github.com/AGMoller/noisy_text

Task

Informative



John Summer
@JohnSummer

Oklahoma's first confirmed case of coronavirus is in Tulsa County
<URL>[#SmartNews](#)

12:00 PM · Jun 1, 2020

23 Retweets 122 Likes



Uninformative



Pete Hanson
@PeterHanson1

Trump could cure Coronavirus 19, AIDS, and Cancer in the same day and the media would say he wasn't doing anything.

12:00 PM · Mar 12, 2020

4 Retweets 29 Likes



Models

- SVM (word 1-2 and char 5-6 grams)
- Neural Networks (200d Twitter GloVe embeddings)
 - Multilayer Perceptron (MLP)
 - Convolutional Neural Network (CNN)
- Transformers
 - BERT-base
 - RoBERTa-base
 - CT-BERT
- Ensemble Methods
 - Hard Voting – Thresholding
 - Soft Voting – Stacking

Test Results

Model	F1
Ensemble (Stacking)	90.54
CT-BERT	90.96
Highest (team NutCracker)	90.96

Evaluation

How do the three types of models perform individually in this classification task?

Unsurprisingly, transformer-based models outperform traditional methods (SVM, feedforward NN).

How important is task-specific pretraining for detection of informative COVID-19 tweets?

Task-specific data is very important. Training on domain-specific CT-BERT (fine-tuned on COVID tweets) led to a 3% higher score as compared to BERT-base and RoBERTa.

Can we ensemble SVM, neural network and BERT-based models to improve robustness?

Our ensemble methods did not beat stand-alone CT-BERT.

Results

