

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

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	Models	Test Results								
<div>Informative</div> <div><div><div><div><div></div><div>John Summer</div><div>@JohnSummer</div></div><div><div>12:00 PM · Jun 1, 2020</div><div>23 Retweets 122 Likes</div></div></div><div><div>Oklahoma's first confirmed case of coronavirus is in Tulsa County</div><div><URL>#SmartNews</div></div></div></div> <div>Uninformative</div> <div><div><div><div></div><div>Pete Hanson</div><div>@PeterHanson1</div></div><div><div>12:00 PM · Mar 12, 2020</div><div>4 Retweets 29 Likes</div></div></div><div><div>Trump could cure Coronavirus 19, AIDS, and Cancer in the same day and the media would say he wasn't doing anything.</div></div></div>	<ul style="list-style-type: none">•SVM (word 1-2 and char 5-6 grams)•Neural Networks (200d Twitter GloVe embeddings)<ul style="list-style-type: none">–Multilayer Perceptron (MLP)–Convolutional Neural Network (CNN)•Transformers<ul style="list-style-type: none">–BERT-base–RoBERTa-base–CT-BERT•Ensemble Methods<ul style="list-style-type: none">– Hard Voting – Thresholding– Soft Voting – Stacking	<table><tr><th>Model</th><th>F1</th></tr><tr><td>Ensemble (Stacking)</td><td>90.54</td></tr><tr><td>CT-BERT</td><td>90.96</td></tr><tr><td>Highest (team NutCracker)</td><td>90.96</td></tr></table> <div>Evaluation</div> <div><div>How do the three types of models perform individually in this classification task?</div><div>Unsurprisingly, transformer-based models outperform traditional methods (SVM, feedforward NN).</div></div> <div><div>How important is task-specific pretraining for detection of informative COVID-19 tweets?</div><div>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.</div></div> <div><div>Can we ensemble SVM, neural network and BERT-based models to improve robustness?</div><div>Our ensemble methods did not beat stand-alone CT-BERT.</div></div>	Model	F1	Ensemble (Stacking)	90.54	CT-BERT	90.96	Highest (team NutCracker)	90.96
Model	F1									
Ensemble (Stacking)	90.54									
CT-BERT	90.96									
Highest (team NutCracker)	90.96									