1. Intorduction

This document describes the file structure that you use in this track. The columns are tab delimited. Your objective it to build a model that can predict "empathy" and "distress" in essays.

The evaluation metric is Pearson correlation with the gold ratings (overall, empathic concern, personal distress). For more information about the data please refer to this paper: <https://www.aclweb.org/anthology/D18-1507/>.

2. The tab delimited file Structure

## message\_id - type text; unique essay id

## response\_id - type text; unique response id

## empathy - type numeric; a real number in range of [1-7]

## distress - type numeric; a real number in range of [1-7]

## empathy\_bin - type numeric; binary [0, 1]

## distress\_bin - type numeric; binary [0, 1]

## emotion\_labels - type text; 6 Ekman basic emotion {joy, sadness, fear, anger, discuss, surprise} & neutral

## essays - type text; collected empathic text from annotator

## demographic information:

### age - type numeric

### gender - type numeric; {Male: 1, Female: 2, Other:5}

### ethnicity - type numberic; {White: 1, Hispanic or Latino: 2; Black or African America: 3, Native American or American Indian: 4, Asian/ Pacific Islander: 5, Other: 6}

### income - type numeric

### education-level - type numeric; {Less than a high school diploma: 1, High school degree or diploma: 2, Technical/Vocational School: 3, Some college â€“ college, university or community college -- but no degree: 4, Two year associate degree from a college, university, or community college: 5, Four year bachelorâ€™s degree from a college or university (e.g., BS, BA, AB): 6, Postgraduate or professional degree, including masterâ€™s, doctorate, medical or law degree (e.g., MA, MS, PhD, MD, JD): 7}

## personality information:

### personality\_conscientiousness - type numeric

### personality\_openess - type numeric

### personality\_extraversion - type numeric

### personality\_agreeableness - type numeric

### personality\_stability - type numeric

### iri\_perspective\_taking - type numeric

### iri\_personal\_distress - type numeric

### iri\_fantasy - type numeric

### iri\_empathatic\_concern - type numeric

3. BibTex to cite this data

@inproceedings{buechel-etal-2018-modeling,

title = "Modeling Empathy and Distress in Reaction to News Stories",

author = "Buechel, Sven and

Buffone, Anneke and

Slaff, Barry and

Ungar, Lyle and

Sedoc, Jo{\~a}o",

booktitle = "Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing",

month = oct # "-" # nov,

year = "2018",

address = "Brussels, Belgium",

publisher = "Association for Computational Linguistics",

url = "https://www.aclweb.org/anthology/D18-1507",

doi = "10.18653/v1/D18-1507",

pages = "4758--4765",

abstract = "Computational detection and understanding of empathy is an important factor in advancing human-computer interaction. Yet to date, text-based empathy prediction has the following major limitations: It underestimates the psychological complexity of the phenomenon, adheres to a weak notion of ground truth where empathic states are ascribed by third parties, and lacks a shared corpus. In contrast, this contribution presents the first publicly available gold standard for empathy prediction. It is constructed using a novel annotation methodology which reliably captures empathy assessments by the writer of a statement using multi-item scales. This is also the first computational work distinguishing between multiple forms of empathy, empathic concern, and personal distress, as recognized throughout psychology. Finally, we present experimental results for three different predictive models, of which a CNN performs the best.",

}