

PROCESSING DATASETS

gathering & preprocessing

sentiment140 (1.6+ million tweets)
emobank
ISEAR
NRC-Emotion-Intensity-Lexicon-v1
-Lexicon-Wordlevel-v0.92
-Hashtag-Emotion-Lexicon-v0.2
NRC-VAD-Lexicon

DATABASE

DATASETS

Preprocessing

Merge & concatenate

merge & split

First split?

Yes

No

Polarities

POSITIVE

NEGATIVE

NLP split into labels

Stanza
SpaCy
Flair
TextBlob
NLTK Vader

Dataframe

Polarities

Labels

NLP split into labels

['anger',
'anticipation',
'surprise',
'trust', 'disgust',
'fear', 'joy',
'sadness']

NLP split into labels

['not confident', 'unsure',
"can't handle pressure", 'disordered',
'talks to much', 'uninterested',
'sad', 'unfriendly', 'confident',
'certain', 'can handles pressure',
'organized', 'concise',
'interested', 'happy', 'friendly']