



Partly Sunny with a Chance of Hashtags

What can a #machine learn from tweets about the #weather?

\$500 · 259 teams · 4 years ago

Competition Data

[Edit](#)[sampleSubmission.csv](#)**sampleSubmission.csv** 2.17 MB[Download](#)[test.csv](#)[train.csv](#)[variableNames.txt](#)

Data Description

[Overview](#)[Data](#)[Discussion](#)[Leaderboard](#)[Rules](#)[Late Submission](#)

come from three categories: sentiment, when, and kind. Human raters can choose only one label from the "sentiment" and "when" categories, but are allowed multiple choices for the "kind". Your goal is to predict a confidence score of all 24 labels for each tweet in the test set.

```
s1,"I can't tell"
s2,"Negative"
s3,"Neutral / author is just sharing information"
s4,"Positive"
s5,"Tweet not related to weather condition"
```

```
w1,"current (same day) weather"
w2,"future (forecast)"
w3,"I can't tell"
w4,"past weather"
```

```
k1,"clouds"
k2,"cold"
k3,"dry"
k4,"hot"
```

```
k5,"humid"  
k6,"hurricane"  
k7,"I can't tell"  
k8,"ice"  
k9,"other"  
k10,"rain"  
k11,"snow"  
k12,"storms"  
k13,"sun"  
k14,"tornado"  
k15,"wind"
```

For example, a tweet "The hot and humid weather yesterday was awesome!" could have $s_4=1$, $w_4=1$, $k_4=1$, $k_5=1$, with the rest marked as zero.

Confidence score?

Each tweet is reviewed by multiple raters, and some amount of disagreement on the labels is expected. The confidence score accounts for two factors - the mixture of labels that the raters gave a tweet and the individual trust of each rater. Since some raters are more accurate than others (e.g. they pay closer attention, take the job more seriously, etc.), these raters count more in the confidence score. You have more confidence that a tweet was referring to the past if you trust the person telling you.

In this competition you do not have access to the individual ratings or the raters' trust. This "unknown trust" issue is therefore a source of noise in the problem. However, you do know that the raters can choose only one label from the "sentiment" and "when" categories, but multiple choices for the "kind". The result is that confidences for the "sentiment" and "when" categories sum to one. Conversely, the sum of the "kind" category will not always sum to one.