



Partly Sunny with a Chance of Hashtags

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

\$500 · 259 teams · 4 years ago

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

Overview

Description

Evaluation

Prizes

Winners



In this competition you are provided a set of tweets related to the weather. The challenge is to analyze the tweet and determine whether it has a positive, negative, or neutral sentiment, whether the weather occurred in the past, present, or future, and what sort of weather the tweet references. It's a lot to mine from so few characters, but if the going gets tough you can always blame the weather...

"Please knock out the power giant storm that is passing thru....please." -Tweet #74096

CrowdFlower

We are excited to team up with CrowdFlower on the first of what we hope will be many fun machine learning projects. CrowdFlower is debuting a new [open data library](#) and we're always looking for an excuse to have a competition. Why is this exciting? Sweet, sweet Labels.

Data repositories sometimes have more in common with a landfill than a library. They're home to tattered piles of spreadsheets in odd formats with nary a shred of documentation to tell the GDP of Chile from the migratory patterns of North American goldfinches. If creating value from this digital exhaust is a defining theme of the big data explosion, most repositories leave you choking on the diesel fumes of data disappointment. Such data is great if you are doing a report on the GDP of Chile, but not so useful if you are doing machine learning, or its red-headed step child, data science.

Crowdfower's data sets provide the thing that makes so many repositories fall short - data paired with labels. One can decide whether two English sentences are related, make

judgments about yogurt chatter, or rank emotions on tweets about nuclear energy. It's all about the (wo)manpower to label what these bytes actually mean.

The Open Data Library

CrowdFlower Open Data Library is a repository of real data set samples that developers, researchers and data scientists can download and use to test and improve algorithms. Our mission is to encourage users to explore the possibilities and power of crowdsourcing. Open Data is free, available to anyone, and ready-to-use with CrowdFlower's [Platform](#).

New data sets are continuously added to CrowdFlower Open Data Library as users of the CrowdFlower Platform opt-in to share their data with the crowdsourcing community. Sample data sets currently available include tweets for sentiment and topic analysis, word combinations to test similarities, sentence combinations to test related topics, and more. Learn more at www.crowdflower.com.

Leaderboard

1

aseveryn

2

Tim Dettmers

3

Duffman & Pascal

4

no_name

5

David

6

orchid

7

JedF

8

ryank

27 discussion topics

Any tips on stacking with multilabels?

4 replies · 9 months ago

Using the data from this competition

6 replies · 2 years ago

Labeled Test Dataset

0 replies · 4 years ago

Congratulations

61 replies · 4 years ago

Removing the cheaters shouldn't affect the inn...

0 replies · 4 years ago

Launch

4 years ago

Close

4 years ago

259

Teams

300

Competitors

Points

This competition awarded standard [ranking points](#)

Tiers

This competition counted towards [tiers](#)

Tags

rmse extra small

