



Featured Prediction Competition

Toxic Comment Classification Challenge

Identify and classify toxic online comments

\$35,000

Prize Money



Jigsaw · 128 teams · 2 months to go (2 months to go until merger deadline)

[Overview](#)[Data](#)[Kernels](#)[Discussion](#)[Leaderboard](#)[Rules](#)[Team](#)[My Submissions](#)[Submit Predictions](#)

Overview

Description

Evaluation

Prizes

Timeline

Discussing things you care about can be difficult. The threat of abuse and harassment online means that many people stop expressing themselves and give up on seeking different opinions. Platforms struggle to effectively facilitate conversations, leading many communities to limit or completely shut down user comments.

The [Conversation AI](#) team, a research initiative founded by [Jigsaw](#) and Google (both a part of Alphabet) are working on tools to help improve online conversation. One area of focus is the study of negative online behaviors, like toxic comments (i.e. comments that are rude, disrespectful or otherwise likely to make someone leave a discussion). So far they've built a range of publicly available models served through the [Perspective API](#), including toxicity. But the current models still make errors, and they don't allow users to select which types of toxicity they're interested in finding (e.g. some platforms may be fine with profanity, but not with other types of toxic content).

In this competition, you're challenged to build a multi-headed model that's capable of detecting different types of toxicity like threats, obscenity, insults, and identity-based hate better than Perspective's [current models](#). You'll be using a dataset of comments from Wikipedia's talk page edits. Improvements to the current model will hopefully help online discussion become more productive and respectful.

Disclaimer: the dataset for this competition contains text that may be considered profane, vulgar, or offensive.



Leaderboard



Kernels



11 discussion topics



1

#!##

2

clustifier

3

Bojan Tunguz

4

Pavel Ostyakov

5

Salil Gautam

6

Mr The Dave

7

extremin

8

Motiur Rahman

NB-SVM baseline and basic E...

38 votes · 6 hours ago

Toxic Regression

18 votes · 10 hours ago

XGB Baseline (0.058 on LB)-c...

3 votes · 4 hours ago

Don't mess with my "mothjer"

4 votes · 8 hours ago

NB-SVM baseline fork

5 votes · 8 hours ago

Natural Language Processing ...

0 replies · 16 minutes ago

Unable to download the dataset

4 replies · an hour ago

Hi! and Welcome to our first to...

0 replies · an hour ago

Kaggle Text related Earlier Co...

0 replies · 2 hours ago

Official External Data Thread

5 replies · 3 hours ago

Launch

16 hours ago

Close

2 months

2 months

Rules Acceptance

Deadline

128

Teams

133

Competitors

Points

This competition awards standard ranking points

Tiers

This competition counts towards tiers

Tags

meancolumnwiselogloss

small