



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

### Competition Data

[Edit](#) [sample\\_submission.cs...](#)**sample\_submission.csv.zip** 1.3 MB [Download](#) [test.csv.zip](#) [train.csv.zip](#)

### Data Description

You are provided with a large number of Wikipedia comments which have been labeled by human raters for toxic behavior. The types of toxicity are:

- `toxic`
- `severe_toxic`
- `obscene`
- `threat`
- `insult`
- `identity_hate`

You must create a model which predicts a probability of each type of toxicity for each comment.

### File descriptions

- **train.csv** - the training set, contains comments with their binary labels
- **test.csv** - the test set, you must predict the toxicity probabilities for these comments. To deter hand labeling, the test set contains some comments which are not included in scoring.
- **sample\_submission.csv** - a sample submission file in the correct format

