

an



## TWITTER SENTIMENT ANALYSIS

Registered

## Knowledge and Learning

## Twitter Sentiment Analysis

In this Practice problem, we present to you some data which has both normal and hate tweets. Your task as a Data Scientist is to identify the tweets which are hate tweets and which are not.

- One person cannot participate with more than one user accounts.
- Appropriate taxes will be applicable on the prize money.

- You are free to use any tool and machine you have rightful access to.
- You can use any programming language or statistical software.

- You are free to use solution checker as many times as you want.
- Adding comment is mandatory for use of solution checker
- Comments will help you to refer to a particular solution at a later point in time.

- At any point in the hackathon, you are expected to respect fellow hackers and act with high integrity.
- Slack Live Chat admins hold the right to blacklist / block any participant found to use foul / disrespectful language. Chat forum will be closely monitored.
- Analytics Vidhya holds the right to disqualify any participant at any stage of competition if found indulged in fraudulent practices.

## Free



## Problem Statement

The objective of this task is to detect hate speech in tweets. For the sake of simplicity, we say a tweet contains hate speech if it has a racist or sexist sentiment associated with it. So, the task is to classify racist or sexist tweets from other tweets.

Formally, given a training sample of tweets and labels, where label '1' denotes the tweet is racist/sexist and label '0' denotes the tweet is not racist/sexist, your objective is to predict the labels on the test dataset.

## Motivation

Hate speech is an unfortunately common occurrence on the Internet. Often social media sites like Facebook and Twitter face the problem of identifying and censoring problematic posts while weighing the right to freedom of speech. The importance of detecting and moderating hate speech is evident from the strong connection between hate speech and actual hate crimes. Early identification of users promoting hate speech could enable outreach programs that attempt to prevent an escalation from speech to action. Sites such as Twitter and Facebook have been seeking to actively combat hate speech. In spite of these reasons, NLP research on hate speech has been very limited, primarily due to the lack of a general definition of hate speech, an analysis of its demographic influences, and an investigation of the most effective features.

## Data

Our overall collection of tweets was split in the ratio of 65:35 into training and testing data. Out of the testing data, 30% is public and the rest is private.

## Data Files

1. **train.csv** - For training the models, we provide a labelled dataset of 31,962 tweets. The dataset is provided in the form of a csv file with each line storing a tweet id, its label and the tweet.  
There is 1 test file (public)
2. **test\_tweets.csv** - The test data file contains only tweet ids and the tweet text with each tweet in a new line.

## Submission Details

The following 3 files are to be uploaded.

1. **test\_predictions.csv** - This should contain the 0/1 label for the tweets in test\_tweets.csv, in the same order corresponding to the tweets in test\_tweets.csv. Each 0/1 label should be in a new line.
2. **A .zip file of source code** - The code should produce the output file submitted and must be properly commented.

## Evaluation Metric:

The metric used for evaluating the performance of classification model would be F1-Score.

The metric can be understood as -

**True Positives (TP)** - These are the correctly predicted positive values which means that the value of actual class is yes and the value of predicted class is also yes.

**True Negatives (TN)** - These are the correctly predicted negative values which means that the value of actual class is no and value of predicted class is also no.

**False Positives (FP)** - When actual class is no and predicted class is yes.

**False Negatives (FN)** - When actual class is yes but predicted class is no.

**Precision** =  $TP / (TP + FP)$

**Recall** =  $TP / (TP + FN)$

**F1 Score** =  $2 * (Recall * Precision) / (Recall + Precision)$

F1 is usually more useful than accuracy, especially if for an uneven class distribution.



📄 Test File (/contest/practice-problem-twitter-sentiment-analysis/download/test-file)

📄 Train File (/contest/practice-problem-twitter-sentiment-analysis/download/train-file)

📄 Sample Submissions (/contest/practice-problem-twitter-sentiment-analysis/download/sample-submission)

## Solution Checker

### Code File

No file chosen

### Solution File\*

(.csv only)

No file chosen

### Solution Description

(max : 180 chars)\*

## Discussions

About the Practice\_Sentiment category (<https://discuss.analyticsvidhya.com/t/about-the-practice-sentiment-category>)

Welcome to Practice Problem : Twitter Sentiment Analysis (<https://discuss.analyticsvidhya.com/t/welcome-to-practice-problem-twitter-sentiment-analysis>)



(<https://analyticsvidhya.com>)

([https://twitter.com/share?utm\\_source=twitter&utm\\_medium=twitter&utm\\_campaign=twitter&utm\\_content=twitter&utm\\_term=twitter](https://twitter.com/share?utm_source=twitter&utm_medium=twitter&utm_campaign=twitter&utm_content=twitter&utm_term=twitter))

UTM\_SOURCE=TWITTER\_CONTEST\_PRACTICE

PROBLEM : TWITTER

SENTIMENT

ANALYSIS&UTM\_MEDIUM=TWITTER&URL=HTTP://DATAHACK.ANALYTICSVIDHYA.COM+/CONTEST/PRACTICE-PROBLEM-TWITTER-SENTIMENT-

ANALYSIS/&TEXT=HEY! I AM

PARTICIPATING IN AN

EXCITING PROBLEM BY

@ANALYTICSVIDHYA

PRACTICE PROBLEM-

TWITTER SENTIMENT

LEARN #DATAHACK

ANALYTICS VIDHYA

About Us (<https://www.analyticsvidhya.com/about-us/>)

Our Team (<https://www.analyticsvidhya.com/team/>)

Career (<https://www.analyticsvidhya.com/career/>)

Contact Us (<https://www.analyticsvidhya.com/contact/>)

Hackathon (<https://datahack.analyticsvidhya.com/hackathon/>)

Discussions (<https://discuss.analyticsvidhya.com/>)

Apply Jobs (<https://www.analyticsvidhya.com/apply-jobs/>)

Leaderboard (<https://datahack.analyticsvidhya.com/leaderboard/>)

Reach Us (<https://datahack.analyticsvidhya.com/reach-us/>)

COMPANIES

Post Jobs (<https://www.analyticsvidhya.com/post-jobs/>)

Trainings (<https://www.analyticsvidhya.com/trainings/>)

Hiring Hackathons (<https://datahack.analyticsvidhya.com/hiring-hackathons/>)

Advertising (<https://www.analyticsvidhya.com/advertising/>)

Reach Us (<https://datahack.analyticsvidhya.com/reach-us/>)

JOIN OUR COMMUNITY :

Facebook (<https://www.facebook.com/analyticsvidhya/>)

Twitter (<https://twitter.com/analyticsvidhya>)

LinkedIn (<https://www.linkedin.com/company/analyticsvidhya/>)

Google Plus (<https://plus.google.com/+AnalyticsVidhya>)

YouTube (<https://www.youtube.com/channel/UC905b92032/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

JOIN OUR COMMUNITY :

Facebook (<https://www.facebook.com/analyticsvidhya/>)

Twitter (<https://twitter.com/analyticsvidhya>)

LinkedIn (<https://www.linkedin.com/company/analyticsvidhya/>)

Google Plus (<https://plus.google.com/+AnalyticsVidhya>)

YouTube (<https://www.youtube.com/channel/UC905b92032/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

Followers (<https://www.linkedin.com/company/analyticsvidhya/>)

