



# ZS Data Scientist Challenge – Summer 2018

Jul 07, 2018, 04:30 AM IST - Jul 23, 2018, 09:30 AM IST

[INSTRUCTIONS](#)[PROBLEMS](#)[SUBMISSIONS](#)[LEADERBOARD](#)[ANALYTICS](#)[JUDGE](#)[← Problems / Predict Sentiments of Tweets](#)

## Predict Sentiments of Tweets

Max. Marks: 1

This problem is no longer available for practice. Apology for any inconvenience!

## Problem Description

Given dataset contains data of tweets on various airline's twitter handles.

It contains a total of 12 columns, out of which one column specifies the sentiment of the tweet. All other columns provide various information related to what was the tweet, where was it posted from, when was it posted, it's retweeted; etc.

Your task is to build a machine learning / deep learning model to predict the sentiment of the tweet using all or some of the other given columns.

[Download Dataset](#)

## Data Description

Description of columns of the dataset is given below -

Sr No	Column name	Description
1.	tweet_id	Id of the tweet
2.	airline_sentiment	Sentiment of the tweet (Target variable)
3.	airline_sentiment_confidence	Confidence with which the given sentiment was determined
4.	negativereason_confidence	Confidence with which the negative reason of tweet was predicted
5.	name	Name of the person who tweeted

6.	retweet_count	Number of retweets
7.	text	Text of the tweet whose sentiment has to be predicted
8.	tweet_created	Time at which the tweet was created
9.	tweet_location	Location from where the tweet was posted
10.	user_timezone	Time zone from where the tweet was posted
11.	negativereason	Reason for which user posted a negative tweet
12.	airline	Airline for which the tweet was posted

Following are the files that will be provided in the dataset -

- train.csv - This file contains all the above-mentioned columns. You are expected to train your models on this file.
- test.csv - This file contains all the above-mentioned columns except “airline\_sentiment” column. You have to predict this column for each record given in this file.
- sample\_submission.csv - This file consists of sample submissions. Your submission should be in exact same format.

## Submission

The submission should be a .csv file stating tweet\_id along with the predicted sentiment of the tweet. Please check the sample\_submissions.csv file and make sure that your submission file in exact same format.

In addition to the predicted events, please provide the following materials in .xls, .xlsx, or .ppt format:

- A 250-word write-up outlining your overall approach to the problem statement
- A list of steps taken in data cleaning, any errors you may have found in the data, and a description of any exclusions in your training dataset
- Details on any pre-processing steps, or any intermediary variables created for the solution
- A summary of any key observations and inferences you made from the data, as well as the most important variables in your model
- A technical explanation of your choice in models, and details on the steps you took to validate your results
- Add this file to the .zip file described below.

For further evaluation, please provide a .zip file containing:

- The code used to create your submission
- A readme document with any information required to reproduce your results from your code 4
- The code should be end-to-end, starting from the extraction of the original files provided to the final submission file.

## Evaluation Metric

The submissions will be evaluated based on F1 Score with 'micro' average. For more information on this metric read [here](#).

[About Us](#)[Innovation Management](#)[Technical Recruitment](#)[University Program](#)[Developers Wiki](#)[Blog](#)[Press](#)[Careers](#)[Reach Us](#)

Site Language: English ▼ | [Terms and Conditions](#) | [Privacy](#) | © 2018 HackerEarth