



Teacher Guide: Global Refugee Crisis

Project Overview

This project is meant to be a relaxed project intended to let students explore. The task is to classify tweets as either having a positive or negative attitude toward refugees; a full scenario is presented in the student document. The project allows students to interact with real Twitter data, but as a consequence, getting high classification accuracies is very difficult. The goal of this project is to allow students to play with real-world data and see firsthand the difficulties of interacting with real, uncleaned data.

Teaching this Project

This is a 3-5 day project. Students can allocate their time as they wish, and should work in groups of 2-3. Begin by briefly introducing the project on the first day (15 minutes, using the intro slides), and then let students explore. This project is not intended to be graded for accuracy -- as long as students put in effort into some aspect of the project, they should receive full credit. Complete directions that students can follow are given in the student guide, which should be clear.



Student Guide: Global Refugee Crisis

The UNHRC, UN Refugee Agency, and Problem Context

In this project, you'll be building the core algorithmic backend of a large project that the United Nations Refugee Agency is working on with Stanford students. Here's an introduction to the full project and its importance.

The UNHCR, the UN Refugee Agency, is a global organization dedicated to saving lives, protecting rights and building a better future for refugees, forcibly displaced communities and stateless people. In response to the current global refugee crisis, a special projects unit called The Hive was recently formed to find creative ways of addressing the global refugee crisis, specifically focusing on American engagement with refugee issues. A critical aspect to The Hive's mission is tracking engagement amongst Americans with regards to refugee-related issues.

Thus, your project will be to develop a social listening tool that can take all the social media activity (on Facebook, Twitter, and Instagram) related to refugees specifically and aggregates the data in such a way that provides us with a clear picture about the current state of American engagement regarding refugee issues.

Your goal will be to extract meaning from tweets and classify them as having a positive or negative attitude toward refugees.





Project Guidelines

This project's goal is to allow you the creative freedom to explore any aspect of the project you find interesting. For a base implementation, you only have to implement one method: **classify**. `Classify` takes in a string, and returns a boolean corresponding to whether the message is positive, or negative.

We won't be telling you how to construct the algorithm, and it'll be incredibly challenging to make a robust classifier. Even consistently hitting 70% by the end of your three days on it is no easy feat. The grading section below explains how you can spend time on the project.

Grading

This project should be **stress free**. The goal of this is to let you explore aspects of the project you find interesting. Of course, you should implement at least a basic `classify` method, but after that, you can spend your time on any of the following:

- Optimizing your algorithm
- Write a learning function to extract information from the training data. Based on characteristics that the computer understands from this data, you can better classify test data.
- Creating a Twitter API key and testing your algorithm against a live stream of Twitter data. Once you do this base implementation, you can query twitter data by geographic location to understand the American refugee climate better.

This project will be graded on the effort you put into it. As long as you're working and exploring aspects of the project you're interested in, you'll receive full credit. If you go above and beyond, there might even be an opportunity for extra credit. **This algorithm is challenging, and you won't be graded on your accuracy.** (Although you might receive bonus points for a great algorithm!)

Going about `classify()`

We're not going to give you any answers to how to classify things, and there's no one right algorithm! That's what makes this project exciting. To build your first classifier, some tips:

- Begin by looking through the data to see if you observe any patterns
- Read into the [split\(\)](#) method of the `String` class to break tweets into words
- Try some basic classifiers, and see how they perform
- To debug your classifier, use `true` as the last parameter (this will show you in red the tweets you classified incorrectly, and in black the ones you got right).

Going about `learn()`

Sometimes, computers can pick up on characteristics of data that you can't. If you're interested, you can have the computer "learn" from a training set and apply that knowledge to a test set. What that learning looks like is up to you. Here are some thoughts:

- Feel free to modify return values, print things out, add variables to the class, etc.
- If interested, look into other Java containers like the [HashMap](#) or `Set`



- Think about what a computer can understand that you can't

Integrating Twitter

To start with, you will need to have a Twitter account and obtain credentials (i.e. API key, API secret, Access token and Access token secret) on the Twitter developer site to access the Twitter API, following these steps:

- Create a Twitter user account if you do not already have one.
- Go to <https://apps.twitter.com/> and log in with your Twitter user account. This step gives you a Twitter dev account under the same name as your user account.
- Click "Create New App"
- Fill out the form, agree to the terms, and click "Create your Twitter application"
- In the next page, click on "Keys and Access Tokens" tab, and copy your "API key" and "API secret". Scroll down and click "Create my access token", and copy your "Access token" and "Access token secret" into the `twitter4j.properties` file in Eclipse.¹

Note: due to school restrictions, you may have to get these keys on a personal computer using a hotspot, or at home.

Now, uncomment the code in `RefugeeCrisis.java` that accesses Twitter data, and see how you do!

For the more adventurous and excited about learning about API's, you can use the Twitter4j API to query tweets from specific geographic locations, or other things that you're interested in. Check out their [documentation](#) and [website](#) if you're interested.

If you're even more excited about using APIs, look into [IBM Watson](#), which can help you extract sentiment from messages. (Note: this will be challenging, but super rewarding!)

¹ Borrowed from: <http://socialmedia-class.org/twittertutorial.html>