# IDMC: Social Media Monitoring

Date: November 27, 2020

Authors: Gokberk Ozsoy, Katharina Boersig, Michaela Wenner, Tabea Donauer

In collaboration with: IDMC

Developed in the context of Hack4Good, 3rd Edition

## Abstract

In recent years, the importance of social media platforms, such as Twitter, in knowledge transfer and information flow has strongly increased. Many individuals as well as organisations and governmental institutions make use of Twitter to efficiently and timely inform followers of happenings around the world. Here, we explore Twitter as a data source for Internal Displacement Monitoring. We use a machine learning (ML) pipeline to filter for relevant tweets and extract important information. 80% of tested tweets are labelled correctly by the classifiers, which gives confidence in its performance. Additionally, a custom trained named entity recognition algorithm (NER) is able to extract the most important information from tested tweets.

## Expected Impact

This project aims at delivering a more complete picture of global internal displacements. Furthermore, more timely information can result in actions being taken faster by governments and non-profit organizations. Potential biases and misinformation of corrupt governments can be overcome by taking reports of individuals and trustworthy institutions into account.

## Approach

Our approach (Fig. 1) consists of two sequential tasks. For the first task, we identify tweets referring to internal displacement. For this, we extract tweets from twitter by filtering for provided keywords of internal displacement. Using a preliminary hard rule decision we make a selection of potentially relevant tweets. Afterwards, we preprocess the tweet's text using different techniques such as lemmatization and stopwords removal. Subsequently, we embed the words via a pretrained word2vec embedding and then classify the tweets using a self-trained binary classifier (SVM). For the second task, we extract information relevant for internal displacement. We apply NER models in order to extract information from the tweet's text. This includes general information such as location and date as well as internal displacement related information such as the unit of displacement and the number of those affected. Finally, this pipeline outputs the relevant tweets by combining the corresponding extracted information and the tweet's important metadata.

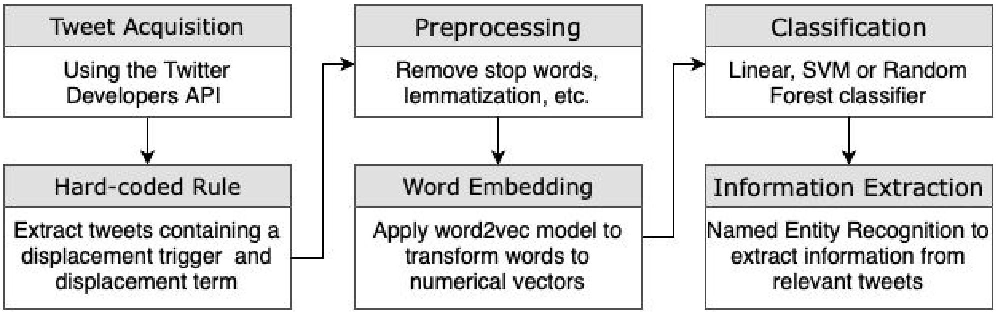

Tweet Acquisition
Using the Twitter Developers API
Hard-coded Rule
Extract tweets containing a displacement trigger and displacement term
Preprocessing
Remove stop words, lemmatization, etc.
Word Embedding
Apply word2vec model to transform words to numerical vectors
Classification
Linear, SVM or Random Forest classifier
Information Extraction
Named Entity Recognition to extract information from relevant tweets

Figure 1: Simplified visualization of the ML pipeline developed for tweet acquisition, classification and information extraction.

## Difficulties, Limitations & Risks

The main limitations of the tool are caused by a) the nature of Twitter as an information platform and b) the available manually labelled data set.

a) Information found on Twitter may be incomplete and often lacks verification. The platform is not equally popular around the world, and the available data may have a spatial as well as a language bias (Fig. 2).

b) Training and validation of our tool depend on manually labelled tweets. In the course of the project we learned how difficult and subtle this decision can be and that the labels have a subjective component. Due to the limited availability of training and validation data it is not clear, whether the proposed algorithm will perform as well on any set of displacement related tweets.

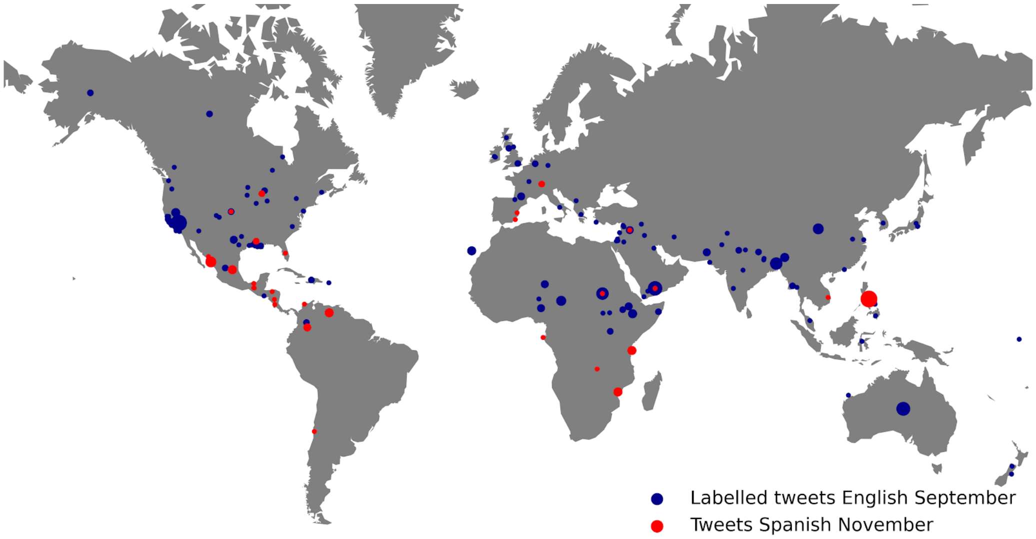
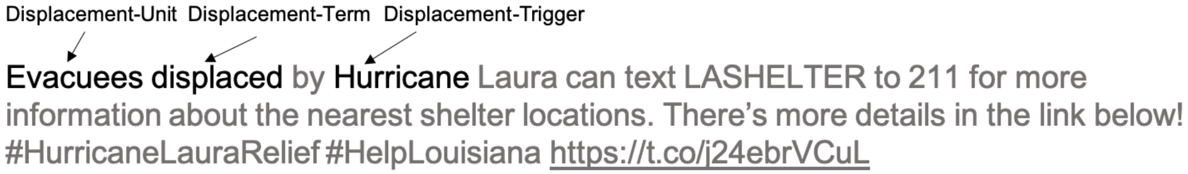

● Labelled tweets English September
● Tweets Spanish November

Figure 2: The classifier has been trained on English displacement-tweets, covering different regions than tweets downloaded in Spanish. Tweet topics are also region- and language-dependent, e.g. lots of the Spanish tweets related to conflicts, while the training dataset had a focus on natural disaster events.

## Results & Deliverables

The developed ML pipeline can be separated in two different main tasks: a) classification in relevant and non-relevant tweets and b) important information extraction. For the classification step, we used 80% of the provided 620 labelled tweets to train a ML classifier (SVMs). To test its accuracy we used the remaining 20%, and compared the predicted labels to the manually assigned ones. We achieve an overall accuracy of about 80%, meaning that 80% of the test data was labelled correctly by the ML classifier.

The custom trained entity recognition algorithm is able to automatically extract the for displacement monitoring important information from never before seen tweets. An example is given in the following:



Evacuees displaced by Hurricane Laura can text LASHELTER to 211 for more information about the nearest shelter locations. There's more details in the link below! #HurricaneLauraRelief #HelpLouisiana https://t.co/j24ebrVCuL

Figure 3: Example of the NER-performance. Input: random tweet, not seen by the NER-algorithm (grey). Output: Named entities (unit, term and trigger) assigned by the NER-algorithm (black).

## Recommendations & Conclusion

The classification accuracy of 80% gives promising results for using Twitter as an information source for internal displacement monitoring. However, the limited amount of training data available during the project prevented a confident prognosis of how the classifier and entity extraction will perform in a "real-life" application. To ensure a satisfying performance we strongly suggest extending the training data set by hundreds of tweets, which cover diverse topics and are ideally downloaded in different languages.