

# Finding event location based on the user generated tweets

## 1 Introduction

Social media platforms have been one of the most informative tools to introduce the events happening worldwide in the past decade. People usually talk about the event before, during, and slightly after each event on their social media. Detecting the event location and the groups connected to the events is very important to many companies whose services are matched with the events. In this project, we want to predict the event's location at the city or country level based on the tweets that have geo\_tags and the tweet's text content. It is a challenging task as people worldwide can talk about every event happening around the globe. So, the location data coming from the tweeter can be very noisy [4]. Although there are algorithms to detect the location of the users in tweeter [5], we assume that the geotag will suffice for the first phase of this project. We will also monitor the text of the tweets in phase two of this project to classify them by their predicted location. For the event that can happen in multiple locations, it will be a multiclass classification problem like [1].

Furthermore, we can find the location of other people who are very interested in the event. We will also investigate the hypothesis that some of the events in different locations may have connections, and we expect to find these connections.

## 2 Dataset

First, we try to find some events happening right now, like Persian Norooz. Then we use tweepy to collect and analyze the tweets using the hashtags related to each event. Twitter allows users to declare their locations, and we will focus on the tweets with location tags (coordinate/place). In the past few years, Tweeter added restrictions on the account's location data. However, as we are more interested in the country or city level location, we should not have difficulties collecting data [2]. If we cannot find enough labeled data, we will manually label the data like [3]. We will need a proper set of train and test data for the classification task. As the geotags in the tweets are very noisy and rare, we may manually label some data sets.

## 3 Evaluation Method

We may have a binary or multiclass classification based on the type of event. The best evaluation methods for this project are Precision, Recall, and F1 scores on the test dataset.

## References

- [1] Bo Han, Afshin Rahimi, Leon Derczynski, and Timothy Baldwin. Twitter geolocation prediction shared task of the 2016 workshop on noisy user-generated text. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, pages 213–217, 2016.
- [2] Anna Kruspe, Matthias Häberle, Eike J Hoffmann, Samyo Rode-Hasinger, Karam Abdulahad, and Xiao Xiang Zhu. Changes in twitter geolocations: Insights and suggestions for future usage. *arXiv preprint arXiv:2108.12251*, 2021.

- [3] Rui Li, Kin Hou Lei, Ravi Khadiwala, and Kevin Chen-Chuan Chang. Tedas: A twitter-based event detection and analysis system. In *2012 IEEE 28Th international conference on data engineering*, pages 1273–1276. IEEE, 2012.
- [4] Hamdy Mubarak and Sabit Hassan. UI2c: Mapping user locations to countries on arabic twitter. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 145–153, 2021.
- [5] Shahab Saquib Sohail, Mohammad Muzammil Khan, and M Afshar Alam. An analysis of twitter users from the perspective of their behavior, language, region and development indices—a study of 80 million tweets. *arXiv preprint arXiv:2105.10245*, 2021.