## Motivation

For what purpose was the dataset created?

## Composition

What do the instances that make up the dataset represent?

Assuming the instances are a sample of larger set, is the sample representative? Why (not)?

Are there any errors, sources of noise, or redundancies in the dataset?

If the dataset is related to people: Can people be identified directly or indirectly from the dataset? Does it contain any sensitive data related to people?

## Collection

How was the data acquired? What mechanism or procedure was used? What was the sampling strategy?

If the dataset is related to people: Were the people notified about data collection? Did they consent to data collection? Were the people provided with a mechanism for revoking their consent? Has an analysis of the potential impact of the data on people been conducted?

## Preprocessing/cleaning/labelling

Has the data been processed / cleaned or labelled? Describe how.