Data Mining - Report

James Donohue (16024143)

Abstract

This is the abstract.
Exercise choices: classification and text mining.

Contents

1 Data Set

The data set chosen for the report is an archive of tweets from the period X.

1.1 Description

Table 1 summarises the attributes in the data set.

Table 1: Summary of data attributes

Attribute	Type	Example value
id	numeric, countably infinite	934592567613755400
lang	nominal, nullable	en
timestamp_ms	numeric, countably infinite	1511659206666

1.1.1 `lang`

The official documentation states of this attribute:

> When present, indicates a BCP 47 language identifier corresponding to the machine-detected language of the Tweet text, or und if no language could be detected. (Twitter, 2018)

References

Twitter (2018) *Tweet data dictionaries*, Available at: https://developer.twitter.com/en/docs/tweets/data-dictionary/overview/tweet-object (Accessed: 8 April 2018).