

Data Mining - Report

James Donohue (16024143)

Abstract

This is the abstract.
Exercise choices: classification and clustering

Contents

1	Data Set	1
1.1	Description	1
	References	2

1 Data Set

The open data set used in this report is the 2016 release of Road Safety Data (Department for Transport, 2017), which gives “the circumstances of personal injury road accidents ... the types (including Make and Model) of vehicles involved and the consequential casualties”. The data is published annually on the data.gov.uk website under a licence that permits non-commercial exploitation (The National Archives, 2017).

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

I will cite Han (2012).

References

Department for Transport (2017) *Road Safety Data - 2016*. <https://data.gov.uk/dataset/cb7ae6fo-4be6-4935-9277-47e5ce24a11f/road-safety-data> Accessed 29 April 2018.

Han, J., Kamber, M. and Pei, J. (2012) *Data Mining: Concepts and Techniques*. 3rd edition, Morgan Kaufmann.

The National Archives (2017) *Open Government Licence v3.0*. <http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/> Accessed 29 April 2018.