**DATA SET DESCRIPTION**

**Context**

The SMS Spam Collection is a set of SMS tagged messages that have been collected for SMS Spam research. It contains one set of SMS messages in English of 5,171 messages, tagged according being ham (legitimate) or spam.

**Content**

· Untitled: Notation of ham or spam

· Text: Sample mails

· Label\_num: class variable (0 or 1)

Number of instances: 5171

Number of attributes: 2 + class attribute

Missing Attribute Values: No

Class Distribution : class value 1 is interpreted as “SPAM” and class value 0 is interpreted as “NOT SPAM”