

Spam E-mail Data

Description

The data consist of 4601 email items, of which 1813 items were identified as spam.

Usage

```
spam7
```

Format

This data frame contains the following columns:

crl.tot

total length of words in capitals

dollar

number of occurrences of the \\$ symbol

bang

number of occurrences of the ! symbol

money

number of occurrences of the word ‘money’

n000

number of occurrences of the string ‘000’

make

number of occurrences of the word ‘make’

yesno

outcome variable, a factor with levels n not spam, y spam

Source

George Forman, Hewlett-Packard Laboratories

These data are available from the University of California at Irvine Repository of Machine Learning Databases and Domain Theories. The address is: <http://www.ics.uci.edu/~Here>

Examples

```
require(rpart)
spam.rpart <- rpart(formula = yesno ~ crl.tot + dollar + bang +
  money + n000 + make, data=spam7)
plot(spam.rpart)
text(spam.rpart)
```