

# Data Mining Assignment 1

Identify a problem from your own experience that you think would be amenable to data mining. For that problem describe:

1. What the data is.
  2. What type of benefit you might hope to get from data mining.
  3. What type of data mining (classification, clustering, etc.) you think would be relevant.
  4. Name one type of data mining that you think would not be relevant, and describe briefly why not.
- For each, illustrate with an example, e.g., if you think clustering is relevant, describe what you think a likely cluster might contain and what the real-world meaning would be.

Write one to two pages of 11 point single-spaced typeset text - you aren't writing a paper, but it isn't short answer either.

Answers:

The problem: classifying whether the email received is Spam or not

- 1) The data will be the words and links present in the email.
- 2) Data mining helps to find the email received is Spam or not Spam
- 3)classification is the best fit for this type of problem
- 4)we can use random forest for this problems