**Classification**

The different classifiers used would each in turn be required to assort each tweet it receives into a category. In this work, the classifiers are always binary, meaning they always assign each tweet into one out of two categories.

Since the purpose of the classifiers is to distill out all useful information and disregard all else, the classes are to be named “News” and “Spam”, where the former would represent all *useful* data and the latter everything, which doesn’t fall into the first category. For the purpose of this paper, information would be regarded as *useful* when it presents knowledge about new events, discoveries or developments which at that point of time are not common knowledge and present added value to their respective readers. These data do not include tweets of obvious commercial or marketing nature. These are to be handled as “Spam”.

The usage of only two classes is made out simplicity consideration. When proven useful the described methods could be further extrapolated to classify the incoming data into further categories. Several other such subclasses to be potentially implemented are “News - Events”, “News – New products” and “Events – Announcements”. The example of obtaining event-relevant information based on a tweeter data feed was demonstrated before **[CITE]**.