**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”.

**Multiclass Classification**

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]**.

**[placeholder] – explain about the different multiclass methods:**

**1. Transform to binary**

**2. Extend from binary**

**3. Hierarchical classfication**

Having several classes of tweets carries with it the benefit of being able to further extract added-value from the Tweeter-feed. Useful data could be segmented into relevant subcategories according to their topics such as technology, politics, business etc. Furthermore, clustering these tweets would allow for summarizing the data and producing concise reports about new developments in their respective fields **[CITE]**. Similar to a news-feed which is already implemented on numerous platforms **[CITE]**.

Breaking down non-news posts into categories might also prove to be advantageous for several purposes. One such use, would be discovering bot-networks. Often bots tend to be interconnected on social networks through following one another or by being in a friendship. This is done to add credibility to bot accounts as well as try and induce Herd-Behavior among other users. A large number of social connections could convince users that a bot is an actual person and influential at that.

Albeit promising, multiclass classification is not without its drawbacks. In order to achieve the additional functionality mentioned above, the classifier must be trained on an appropriate training set. This training said should at least **n**-times larger than a binary classifier, where **n** is the number of different class featured in the multiclass. This additional dataset size should provide more examples of each particular class for the algorithm to learn from.

**Clickbait**

Another common type or *class* ofTweets, which are commonly referred to as *Clickbait*, are quite wide spread when concentrating on the e-Commerce theme. According to the Merriam-Webster Dictionary: *Clickbait* *is something (such as a headline) designed to make readers want to*[*click*](https://www.merriam-webster.com/dictionary/click)*on a hyperlink especially when the link leads to content of dubious value or interest.* In the context of Twitter, Clickbait messages are camouflaged to seem as genuine news content, which entice the user to follow a link where further information is to be revealed. However, upon following the link, it becomes obvious that the information presented is obviously online advertising and the purpose of the Tweet is to generate income for its poster through clicks. Many online advertising platforms have a payout schemes, which reward clicks of views of some certain webpages. *Clickbait* exploit such schemes by posting Tweets with [sensationalist](https://en.wikipedia.org/wiki/Sensationalism) [headlines](https://en.wikipedia.org/wiki/Headline) or eye-catching [thumbnail](https://en.wikipedia.org/wiki/Thumbnail) pictures to attract [click-throughs](https://en.wikipedia.org/wiki/Click-through).

This particular class of Tweet would be hard to distinguish from openly obvious advertising, since it is exactly meant to mimic the veneer of proper news posts.

**Hashtag Spamming**

---placeholder--- using trending hashtags in non-related messages. Usually spam