**Data Preparation**

After examining the structure of the Tweet dataset, the *tweet* field that was of factor type was converted to character data. Since the *id* variable had no meaning for the analysis, that column was removed. Since it is free text, the tweet variables were altered further by converting the characters to lower case, removing English stop words, deleting non-alphabetic characters, and shortening all URLs to simply https or http. This URL shortening was performed as the researchers thought that the fact that URLs were including could be important for the modeling and analysis being performed.