

The target variable for this sentiment analysis project is the "emotion type", which represents the sentiment expressed in each tweet. The goal is to classify tweets into four categories: positive, negative, irrelevant, or neutral, making "emotion type" the key variable we want to predict in order to understand public sentiment. The predictors for this project are primarily the text of the tweets, found in the "tweet text", which is preprocessed to ensure effective model training. Other features like "information" can also be used as predictors, as they provide context that might influence the sentiment expressed. In terms of dataset exploration, the dataset consists of four columns: company\_id, information, emotion type, and tweet text. The dataset contains several rows, with each row representing a tweet, and these features are of type string, except for the sentiment labels, which are categorical. The overall dataset is then explored to determine unique values and check for missing or duplicated entries, ensuring that the final cleaned dataset is ready for sentiment classification.