

Codebook: Using Text-Mining Methods to Understand Guest Speaker Presentations in Statistics 141SL

Nikhil Sharma, Tiffany Trinh, Huiyu Chuang, Eddie Liu, Nicholas Ortega

Variable	Description
Comment	Character: Raw, unedited response; provided by Professor
Speaker	Categorical: Speaker to whom response corresponds to; provided by professor
Tag	Binary: Indicated whether a response is thorough (1) or not (0); tagged manually by group
Word Count	Numerical: Amount of words in response; found via Python code
Negative Score	Numerical: Numerical measure of negativity in sentiment of response; calculated via VADER Sentiment Analysis in Python
Neutral Score	Numerical: Numerical measure of neutrality in sentiment of response; calculated via VADER Sentiment Analysis in Python
Positive Score	Numerical: Numerical measure of positivity in sentiment of response; calculated via VADER Sentiment Analysis in Python
Compound Score	Numerical: Weighted, normalized measure of Negative, Neutral and Positive Scores; calculated via VADER Sentiment Analysis in Python
Cleaned Comment	Character: response with words lemmatized and stop words removed; found via Python code

Frequency of Top Unigrams	Numerical: Number of times the top 30 unigrams by speaker appeared in the response; found via R code
Frequency of Top Bigrams	Numerical: Number of times the top 30 bigrams by speaker appeared in the response; found via R code