**Agreement of sentiment analysis tools with the manual labeling and with each other**

|  |  |  |
| --- | --- | --- |
| **TOOLS** | **KAPPA** | **ARI** |
| NLTK​ ​ vs​ ​ Manual | 0.3 | 0.06 |
| SentiStrength​ ​ vs​ ​ Manual | 0.39 | 0.139 |
| WatsonNLU​ ​ vs​ ​ Manual | 0.43 | 0.177 |
| TextAnalytics vs Manual | 0.4 | 0.112 |
| NLTK​ ​ vs​ ​ SentiStrength | 0.3 | 0.069 |
| NLTK vs WatsonNLU | 0.49 | 0.221 |
| NLTK vs TextAnalytics | 0.38 | 0.119 |
| SentiStrength vs WatsonNLU | 0.49 | 0.202 |
| SentiStrength vs TextAnalytics | 0.31 | 0.07 |
| WatsonNLU vs TextAnalytics | 0.36 | 0.114 |

**Agreement of groups of tools with the manual labeling (n = the number of comments the tools agree upon)**

|  |  |  |  |
| --- | --- | --- | --- |
| **TOOLS** | **n** | **KAPPA** | **ARI** |
| NLTK​, SentiStrength | 188 | 0.6 | 0.297 |
| NLTK, WatsonNLU | 242 | 0.51 | 0.24 |
| NLTK, TextAnalytics | 208 | 0.51 | 0.184 |
| SentiStrength, WatsonNLU | 238 | 0.55 | 0.283 |
| SentiStrength, TextAnalytics | 182 | 0.64 | 0.377 |
| WatsonNLU, TextAnalytics | 195 | 0.65 | 0.381 |
| NLTK​, SentiStrength, WatsonNLU | 159 | 0.66 | 0.405 |
| NLTK​, SentiStrength, TextAnalytics | 121 | 0.77 | 0.554 |
| NLTK​, WatsonNLU, TextAnalytics | 147 | 0.73 | 0.487 |
| SentiStrength, WNLU, TextAnalytics | 139 | 0.73 | 0.519 |
| NLTK, SS, WNLU, TextAnalytics | 109 | 0.81 | 0.661 |