|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Trait** | **Ratio***c* **of Present to Absent** | **Gain - Correct Prediction of Presence** | **Cost - Wrong Prediction of Presence** | **Gain - Correct Prediction of Absence** | **Cost - Wrong Prediction of Absence** |
| Actionable | 677 : 1248 | +1.84 | -1.00 | +1.00 | -1.84 |
| Behavior Focused | 1093 : 832 | +1.00 | -1.310 | +1.310 | -1.00 |
| Detailed | 629 : 1296 | +2.06 | -1.00 | +1.00 | -2.06 |
| Negative Feedback | 512 : 1413 | +2.76 | -1.00 | +1.00 | -2.76 |
| Professional Communication | 467 : 1458 | +3.12 | -1.00 | +1.00 | -3.12 |
| Specific | 704 : 1221 | +1.73 | -1.00 | +1.00 | +1.73 |

**Table, Supplementary Digital Content 6**. Details on Cost Matrices for Trait Models’ Development

*a*“Presence” refers to a trait having been identified as present by the original raters. “Absence” refers to a trait having not been identified as present by the original raters.

*b*The positive class was set as “trait present” for these models. Therefore, “Correct Prediction of Presence” is analogous to a “True Positive,” “Wrong Prediction of Presence” is analogous to a “False Positive,” “Correct Prediction of Absence” is analogous to a “True Negative,” and “Wrong Prediction of Absence” is analogous to a “False Negative.”

*c*Ratio represents the number of comments scored by the original raters as having the trait present compared to the number of comments scored as having the trait absent. After dividing the number of comments for the predominant class of each trait over the minority class, the resultant ratio was used to set the gain and cost values for the minority class. With the majority class’s gain and cost values set as +1 and -1 respectively, this results in a cost matrix which encourages development of models that correctly predict the minority class instead of simply favoring the majority class, based on the ratio of classes in the original dataset.