Appendix D. Out-of-sample performance per fine-tuned BERT classifier

Classifier	Accuracy	Weighted F1	Support
1A	0.977	0.978	300
1B	0.977	0.977	300
2A	0.953	0.953	300
2B	0.937	0.932	300
2C	0.903	0.892	300
3A	0.980	0.979	300
4A	0.860	0.850	300

Note. The names in the format 1A-4B represent the different fine-tuned BERT classifiers. Details of how those were trained can be found in §2.2.5. Support in means number of observations in this category on the out-of-training-sample data. F1 is the harmonic mean of recall and precision.