APPENDIX S5 for

"Following the blind? Database Policies and the Case of IFRS Noncompliance"

This supplementary appendix describes the samples of English, French, and German language reports used to fine-tune pre-trained Bidirectional Encoder Representations from Transformers (BERT) models for classifying the consolidation status of financial statements. It also provides statistics on the BERT base model's consolidation classifications compared to the Worldscope data item "Accounting Method For Long Term Investment>50%" (WC07531) for our sample of 43,069 English-language reports in the EU from 2000 to 2018.

On Huggingface, we provide researchers with our BERT training datasets, the BERT base model's predicted consolidation classification, and an identifier file with links to Worldscope.[1] On GitHub, we offer a customizable Python code that can be used for similar classification tasks.[2]

Table of Contents:

TABLE S5.1 BERT training sample for English, French, and German language annual reports

Period	English		French		German	
	N	%	N	%	N	%
2000	27	1.16	29	0.57	42	0.53
2001	71	3.05	285	5.60	366	4.66
2002	97	4.17	316	6.21	502	6.39
2003	131	5.63	355	6.97	527	6.71
2004	117	5.03	376	7.39	540	6.88
2005	200	8.60	450	8.84	668	8.51
2006	288	12.38	475	9.33	728	9.27
2007	209	8.99	482	9.47	801	10.20
2008	305	13.11	448	8.80	755	9.62
2009	190	8.17	442	8.68	754	9.60
2010	285	12.25	453	8.90	723	9.21
2011	127	5.46	464	9.12	690	8.79
2012	141	6.06	513	10.08	749	9.54
2013	138	5.93	2	0.04	6	0.08
Total	2,326		5,090		7,851	

Notes: This table presents the distribution of BERT training data by year. For each of our training datasets, we present the total number of annual reports and the percentage for each year.

---

[1] https://huggingface.co/Databasesprojec
[2] https://github.com/Databasesprojec?tab=repositories

TABLE S5.2 Consolidation classification for BERT base model vs. Worldscope

Period	N	Consolidated	
		BERT base (%)	Worldscope (%)
2000	1,179	94.06	98.13
2001	1,708	93.03	97.78
2002	1,838	92.98	97.66
2003	1,992	93.72	97.79
2004	2,268	93.03	98.06
2005	2,505	92.65	97.56
2006	2,634	92.48	97.84
2007	2,539	92.79	97.99
2008	2,451	93.10	98.08
2009	2,442	92.75	98.24
2010	2,415	92.71	98.30
2011	2,330	91.76	98.67
2012	2,344	92.36	99.40
2013	2,440	91.60	99.51
2014	2,445	91.78	99.75
2015	2,483	92.71	99.72
2016	2,418	93.13	99.75
2017	2,498	92.59	99.92
2018	2,140	92.38	99.95
Total	43,069		

Notes: This table presents a comparison of their consolidation classifications from 2000–2018 between the fine-tuned BERT base model and the Worldscope data item "Accounting Method For Long Term Investment>50%" (WC07531). We present the yearly total number of observations and percentages of annual reports classified as "consolidated".