

Appendix A

Table 1: Dev set F1 scores (%) of the machine learning models

SBERT before SMOTE			SBERT after SMOTE	
	APC	ATM	APC	ATM
Random forest	77.7 $\pm$ 0.11	71.5 $\pm$ 0.72	48.8 $\pm$ 24.4	74.5 $\pm$ 2.65
XGBoost	74.2 $\pm$ 0.29	78.2 $\pm$ 0.09	74.2 $\pm$ 0.27	78.2 $\pm$ 0.09
LightGBM	77.1 $\pm$ 0.20	74.8 $\pm$ 0.64	77.1 $\pm$ 0.21	74.8 $\pm$ 0.64

SimCSE before SMOTE			SimCSE after SMOTE	
	APC	ATM	APC	ATM
Random forest	77.6 $\pm$ 0.11	77.9 $\pm$ 0.20	48.2 $\pm$ 24.1	75.1 $\pm$ 2.89
XGBoost	74.2 $\pm$ 0.46	79.1 $\pm$ 0.14	74.2 $\pm$ 0.46	72.4 $\pm$ 0.68
LightGBM	76.9 $\pm$ 0.23	79.5 $\pm$ 0.13	76.9 $\pm$ 0.23	75.4 $\pm$ 0.29

Table 2: Test set F1 scores (%) of the machine learning models

SBERT before SMOTE			SBERT after SMOTE	
	APC	ATM	APC	ATM
Random forest	78.2 $\pm$ 0.27	78.2 $\pm$ 0.19	78.2 $\pm$ 0.31	78.2 $\pm$ 0.18
XGBoost	79.2 $\pm$ 0.29	78.5 $\pm$ 0.19	79.2 $\pm$ 0.29	78.5 $\pm$ 0.19
LightGBM	79.7 $\pm$ 0.31	78.8 $\pm$ 0.19	79.7 $\pm$ 0.31	78.8 $\pm$ 0.19

SimCSE before SMOTE			SimCSE after SMOTE	
	APC	ATM	APC	ATM
Random forest	78.2 $\pm$ 0.30	78.6 $\pm$ 0.17	78.2 $\pm$ 0.31	78.6 $\pm$ 0.15
XGBoost	79.3 $\pm$ 0.29	79.4 $\pm$ 0.11	79.3 $\pm$ 0.29	79.4 $\pm$ 0.11
LightGBM	79.7 $\pm$ 0.31	79.7 $\pm$ 0.22	79.7 $\pm$ 0.31	79.7 $\pm$ 0.22

Table 3: Dev set Recall scores (%) of the machine learning models

SBERT before SMOTE			SBERT after SMOTE	
	APC	ATM	APC	ATM
Random forest	91.8 $\pm$ 0.37	76.1 $\pm$ 1.30	43.4 $\pm$ 21.8	84.7 $\pm$ 2.27
XGBoost	83.6 $\pm$ 0.82	91.4 $\pm$ 0.05	83.6 $\pm$ 0.82	91.4 $\pm$ 0.05
LightGBM	91.8 $\pm$ 0.45	83.8 $\pm$ 1.35	91.8 $\pm$ 0.45	83.8 $\pm$ 1.35

SimCSE before SMOTE			SimCSE after SMOTE	
	APC	ATM	APC	ATM
Random forest	91.8 $\pm$ 0.35	89.8 $\pm$ 0.39	42.7 $\pm$ 21.4	86.1 $\pm$ 2.18
XGBoost	83.6 $\pm$ 0.94	93.8 $\pm$ 0.35	83.6 $\pm$ 0.94	77.4 $\pm$ 1.02
LightGBM	91.2 $\pm$ 0.59	95.6 $\pm$ 0.14	91.2 $\pm$ 0.59	85.2 $\pm$ 0.70

Table 4: Test set Recall scores (%) of the machine learning models

SBERT before SMOTE			SBERT after SMOTE	
	APC	ATM	APC	ATM
Random forest	93. $\pm$ 0.27	90.5 $\pm$ 0.14	93. $\pm$ 0.34	90.5 $\pm$ 0.10
XGBoost	97.5 $\pm$ 0.22	92.4 $\pm$ 0.17	97.5 $\pm$ 0.22	92.4 $\pm$ 0.17
LightGBM	99.4 $\pm$ 0.08	94. $\pm$ 0.11	99.4 $\pm$ 0.08	94. $\pm$ 0.11

SimCSE before SMOTE			SimCSE after SMOTE	
	APC	ATM	APC	ATM
Random forest	93.1 $\pm$ 0.14	91.6 $\pm$ 0.16	93.1 $\pm$ 0.25	91.5 $\pm$ 0.22
XGBoost	97.6 $\pm$ 0.17	94.8 $\pm$ 0.20	97.6 $\pm$ 0.17	94.8 $\pm$ 0.21
LightGBM	99.3 $\pm$ 0.08	96. $\pm$ 0.21	99.3 $\pm$ 0.08	96. $\pm$ 0.21

Table 5: Dev set Precision scores (%) of the machine learning models

SBERT before SMOTE			SBERT after SMOTE	
	APC	ATM	APC	ATM
Random forest	67.3 $\pm$ 0.08	68.2 $\pm$ 0.50	56.1 $\pm$ 28.1	66.9 $\pm$ 5.24
XGBoost	66.6 $\pm$ 0.27	68.3 $\pm$ 0.15	66.6 $\pm$ 0.27	68.3 $\pm$ 0.15
LightGBM	66.5 $\pm$ 0.20	67.6 $\pm$ 0.52	66.5 $\pm$ 0.20	67.6 $\pm$ 0.52

SimCSE before SMOTE			SimCSE after SMOTE	
	APC	ATM	APC	ATM
Random forest	67.3 $\pm$ 0.06	68.9 $\pm$ 0.16	55.7 $\pm$ 27.9	66.7 $\pm$ 5.36
XGBoost	66.6 $\pm$ 0.16	68.3 $\pm$ 0.19	66.6 $\pm$ 0.16	68. $\pm$ 0.71
LightGBM	66.4 $\pm$ 0.19	68. $\pm$ 0.18	66.4 $\pm$ 0.19	67.7 $\pm$ 0.48

Table 6: Test set Precision scores (%) of the machine learning models

SBERT before SMOTE			SBERT after SMOTE	
	APC	ATM	APC	ATM
Random forest	67.4 $\pm$ 2.14	68.8 $\pm$ 0.25	67.5 $\pm$ 0.44	68.8 $\pm$ 0.21
XGBoost	66.7 $\pm$ 0.47	68.3 $\pm$ 0.25	66.7 $\pm$ 0.47	68.3 $\pm$ 0.25
LightGBM	66.5 $\pm$ 0.45	67.9 $\pm$ 0.24	66.5 $\pm$ 0.45	67.9 $\pm$ 0.24

SimCSE before SMOTE			SimCSE after SMOTE	
	APC	ATM	APC	ATM
Random forest	67.5 $\pm$ 0.48	68.9 $\pm$ 0.20	67.4 $\pm$ 0.45	68.9 $\pm$ 0.24
XGBoost	66.7 $\pm$ 0.49	68.3 $\pm$ 0.21	66.7 $\pm$ 0.49	68.3 $\pm$ 0.22
LightGBM	66.5 $\pm$ 0.45	68.1 $\pm$ 0.24	66.5 $\pm$ 0.45	68.1 $\pm$ 0.25