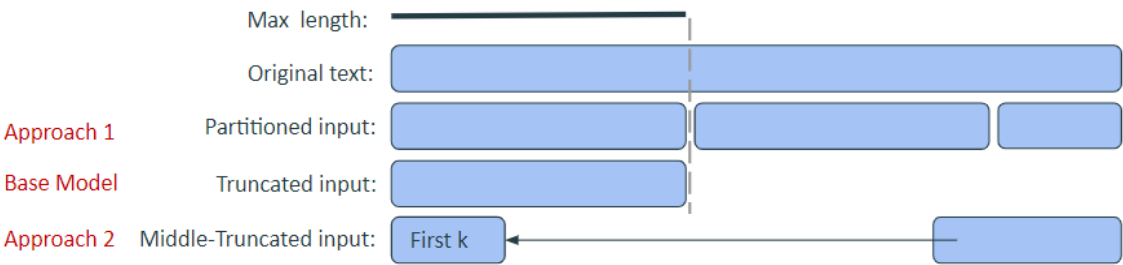
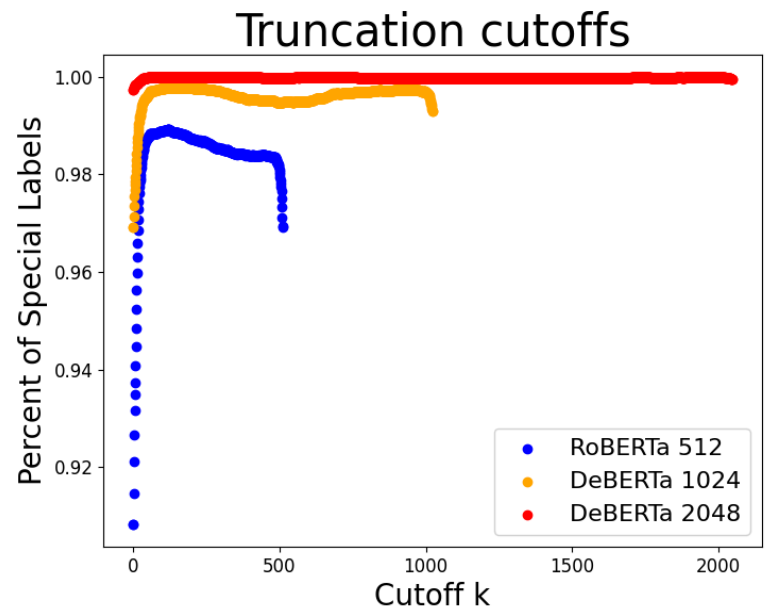


Graphic for Preprocessing Approaches: Partition vs. Truncation



Graph of %PII Found with Various Truncation Approaches



Optimal cutoff for RoBERTa 512 is 121
Optimal cutoff for DeBERTa 1024 is 175
Optimal cutoff for DeBERTa 2048 is 1024

Table of Results

Model (R for Roberta, D for Deberta)	Method (P for partition, T for truncation)	Penalty	Downsampling (Y for yes, N for no)	Epoch 1 (f5 eval)	Epoch 2 (f5 eval)	Epoch 3 (f5 eval)	Epoch 4 (f5 eval)	Epoch 5 (f5 eval)	Epoch 6 (f5 eval)	Private Submission Score (On best epoch)	Public Submission Score (On best epoch)
R	P	.02	N	76.76	82.69	85.94	87.96	91.19	92.84		
R	P	.03	N	83.17	92.22	86.12	88.63	91.20	92.78		
R	P	.04	N	80.70	85.76	86.08	88.05	89.99	93.62		
R	P	.05	N	82.33	87.53	86.33	88.12	93.78	93.76	83.885	87.110
R	P	.06	N	82.24	79.82	84.71	88.68	93.61	93.68		
R	P	.07	N	77.50	81.69	85.82	87.63	88.12	91.49		
R	P	.03	Y	88.39	89.10	90.56	94.49	95.38	95.17		
R	P	.05	Y	90.34	88.12	91.17	95.91	95.63	95.60	84.755	88.578
R	P	.07	Y	88.39	89.10	90.56	94.49	95.38	95.17		
D	P	.05	N	83.15	79.41	88.67	84.64	87.16	91.46	T = .9 83.143 — T = .95 83.216 —	T = .9 86.407 — T = .95 86.460 —
D	P	.05	Y	82.79	88.89	90.89	N/A	N/A	N/A	T=.9 85.697 — T=.99 86.227 — T=.95 85.830	T=.9 88.171 — T=.99 84.481 — T=.95 88.320
D	P	0.05	Y	88.14	86.73	89.78	84.25	92.01	92.17	T = 0.9 85.54 — T = 0.99 85.55	T = 0.9 87.47 — T = 0.99 87.57
D	T	.05	N	82.72	90.85	87.40	88.45	88.54			
D	T	.05	N	84.77	84.92	90.11	95.75	93.17	N/A	T = .9 84.968 — T = .95 85.099 — T = .99 85.356	T = .9 86.581 — T = .95 86.625 — T = .99 86.840

* T indicates the threshold for declaring something a non PII-label

* Bolding indicates epoch for Kaggle scoring

Baseline: 81.715, 84.920

Baseline + Custom Loss Function: 82.842, 85.848

Table of Final Model Scores (Models Trained on All Data)

Model (R or D)	Method (P or T)	Penalty	Downsampling ? (Y or N)	Threshold (T-value)	Private Submission Score	Public Submission Score
D	P	.05	Y	.9	83.793	86.896
D	P	.05	Y	.99	85.537	87.466
D	T	.05	N	.9	84.968	86.581
D	T	.05	N	.99	85.356	86.840