

Model	Convergence	Time		Val loss	Val acc	Performance
BERT	Epoch 2	1 hr 30 min	Less time	2.05	35.5 %	Worst
XLNet	Epoch 3	3 hr	More time	1.50	45.5 %	Better