Max Value:

Max of value												
Row Labels	**EpochTime**	**GPU/Memory(MiB)**	**GPU/Utilization(%)**	**GradientNorm**	**HitRate@10/test**	**LearningRate**	**Loss/train**	**Model/TotalParams**	**MRR@10/test**	**NDCG@10/test**	**Recall@10/test**	**Total Epoch Trained**
BERT4Rec												
100K												
50	6.91	1177	38	7.59	0.14	0.001	4888.11	274835	0.047	0.068	0.136	6
75	6.94	1177	35	7.03	0.13	0.001	4884.57	276435	0.045	0.064	0.131	5
100	6.90	1177	36	7.84	0.12	0.001	4877.71	278035	0.039	0.057	0.118	6
1M												
50	92.87			6.07	0.29	0.001	45807.93	1372923	0.126	0.163	0.286	9
75	93.73			4.85	0.29	0.001	45805.49	1376123	0.127	0.164	0.290	12
100	92.43			4.96	0.29	0.001	45809.74	1379323	0.127	0.164	0.290	13
SAS4Rec												
100K												
50	5.97	1177	0	3.47	0.14	0.001	5113.95	270291	0.049	0.068	0.136	7
75	6.05	1177	35	4.12	0.13	0.001	5144.62	271891	0.046	0.064	0.133	12
100	6.04	1177	33	3.32	0.13	0.001	5112.20	273491	0.042	0.060	0.128	6
1M												
50	83.36			5.13	0.28	0.001	47815.95	1355643	0.125	0.162	0.284	14
75	83.39			2.72	0.28	0.001	48036.96	1358843	0.123	0.160	0.280	14
100	84.83			2.76	0.28	0.001	48267.37	1362043	0.123	0.160	0.282	14
xLSTM												
100K												
50	20.49	1177	38	3.41	0.12	0.001	5142.62	269188	0.045	0.062	0.118	4
75	12.68	1177	40	3.31	0.12	0.001	5156.19	269188	0.040	0.057	0.120	4
100	64.24	1177	40	3.74	0.11	0.001	5150.41	269188	0.042	0.059	0.115	5
1M												
50					0.20		51217.27		0.087	0.113	0.198	9
75					0.21		50743.71		0.094	0.122	0.214	7
100	220.68			1.44	0.22	0.001	50920.25	1376904	0.093	0.122	0.215	10
Grand Total (Max)	**220.68**	**1177**	**40**	**7.84**	**0.29**	**0.001**	**51217.27**	**1379323**	**0.127**	**0.164**	**0.290**	

Testing Scope: (27 Combinations)

Three Models, Three Datasets, Three Sequence Length

Models	Bert4Rec	Sas4Rec	xLSTM	
Datasets	100K	1M	10M	Steam (optionally)
Sequence Length	100	75	50	

All models use a consistent 1177 MiB of GPU memory, likely due to fixed max length and embedding dimensions.

Model Capacity

Model sizes:

BERT4Rec ~1.37M params (on 1M)

xLSTM ~1.37M params (on 1M)

Trade-off Between Performance vs. Efficiency

Model	Pros	Cons
BERT4Rec	Best metrics on 1M (Reca	Extremely high training time (EpochTime ~92s/epoch), model size (~1.3M params)
SAS4Rec	Decent performance with	Slightly lower accuracy than BERT4Rec
xLSTM	Much faster training time	Performance drops, especially on 1M dataset

If training under a time or compute constraint (e.g., for edge deployment, online learning), xLSTM is worth considering despite performance drops.

xLSTM large work best for 10M, 20M, testing in progress.