

Model	Dataset	Seq Len	TotalParams	Hit Rate@10	NDCG@10	MRR@10	Mean_EpochTime
BERT4Rec	100K	32	273,811	0.1273 ± 0.0084	0.0620 ± 0.0044	0.0425 ± 0.0042	7.00205
BERT4Rec		64	275,859	0.1217 ± 0.0091	0.0582 ± 0.0056	0.0392 ± 0.0050	7.03565
BERT4Rec		128	279,955	0.1202 ± 0.0097	0.0584 ± 0.0042	0.0399 ± 0.0035	7.12781
BERT4Rec	1M	32	1,370,875	0.2766 ± 0.0149	0.1573 ± 0.0103	0.1209 ± 0.0088	95.4389
BERT4Rec		64	1,374,971	0.2752 ± 0.0155	0.1550 ± 0.0106	0.1184 ± 0.0090	95.1566
BERT4Rec		128	1,383,163	<b>0.2777 ± 0.0174</b>	<b>0.1568 ± 0.0110</b>	<b>0.1200 ± 0.0092</b>	<b>101.322</b>
SAS4Rec	100K	32	269,139	0.1278 ± 0.0098	0.0623 ± 0.0038	0.0427 ± 0.0028	5.92104
SAS4Rec		64	271,187	0.1247 ± 0.0106	0.0606 ± 0.0059	0.0414 ± 0.0051	6.0031
SAS4Rec		128	275,283	0.1283 ± 0.0106	0.0615 ± 0.0056	0.0415 ± 0.0044	6.0581
SAS4Rec	1M	32	1,353,339	0.2109 ± 0.0120	0.1191 ± 0.0083	0.0912 ± 0.0072	82.2103
SAS4Rec		64	1,357,435	0.1427 ± 0.0072	0.0798 ± 0.0053	0.0606 ± 0.0047	82.361
SAS4Rec		128	1,365,627	0.0835 ± 0.0048	0.0461 ± 0.0032	0.0347 ± 0.0027	89.9583
xLSTM	100K	32	269,188	0.1041 ± 0.0070	0.0509 ± 0.0036	0.0350 ± 0.0029	9.60462
xLSTM		64	269,188	0.1057 ± 0.0083	0.0522 ± 0.0050	0.0361 ± 0.0046	10.0209
xLSTM		128	269,188	0.1036 ± 0.0095	0.0511 ± 0.0044	0.0355 ± 0.0038	10.3313
xLSTM	1M	32	1,376,904	<b>0.2625 ± 0.0146</b>	<b>0.1492 ± 0.0102</b>	<b>0.1148 ± 0.0089</b>	<b>143.17</b>
xLSTM		64	1,376,904	0.2603 ± 0.0151	0.1483 ± 0.0104	0.1143 ± 0.0089	147.064
xLSTM		128	1,376,904	0.2605 ± 0.0155	0.1479 ± 0.0107	0.1136 ± 0.0092	164.203

### Data Splitting Protocol and Experiment

- Training:** All interactions except the final two ( $t = 1$  to  $N - 2$ ).
- Validation Set:** Predict the  $(N - 1)$ th item. [2]
- Test Set:** Predict the final item using the full sequence.
- Sequence Length:** 32, 64, 128
- Total Experiments:** 84 ML Experiments, 3 seeds

### Conclusion

•xLSTM evaluated under a **novel configuration** for Sequential recommenders; observed performance on various conditions.

•**Performance Scaling (RQ1):** xLSTM matches BERT4Rec's Recall@10 (~26-27%) on the 1M dataset, indicating scalability with richer interaction histories. Performance converges as dataset size grows.

•**Sequence Sensitivity (RQ2):** Standard deviation increases with sequence length, underscoring sensitivity to input length variations.

•**Trade-offs (RQ3):** xLSTM achieves competitive accuracy on large datasets but incurs higher computational costs, especially in smaller-scale scenarios.

•**Baseline Robustness:** BERT4Rec consistently outperforms or matches alternatives across dataset scales and configurations.

Mean_HitRate@10/test	Mean_Loss/train
0.127253	4251.71
0.121675	4219.38
0.120245	4233.56
0.276573	39896.5
0.275211	40411.9
0.277696	40175.6
0.127809	4204.12
0.124696	4273.42
0.12827	4145.57
0.210902	40725.6
0.14274	40545.7
0.0834506	40666.8
0.1041	4447.24
0.105735	4441.26
0.103555	4460.42
0.26248	40216.3
0.260296	40450
0.26049	40522.2