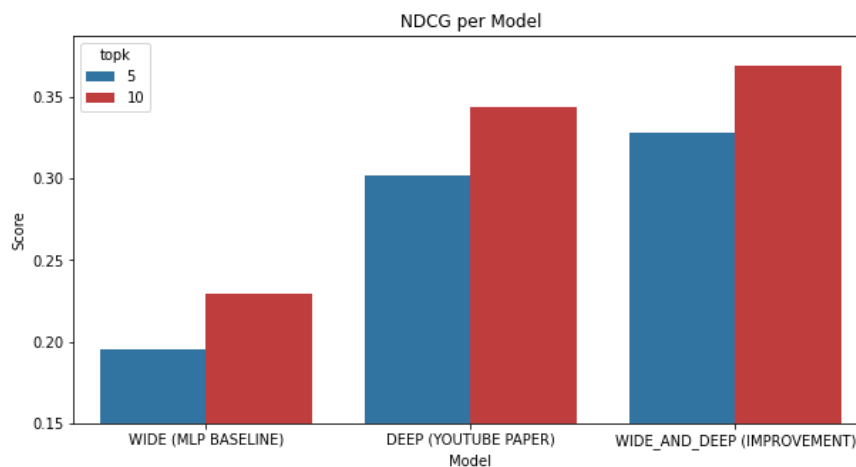
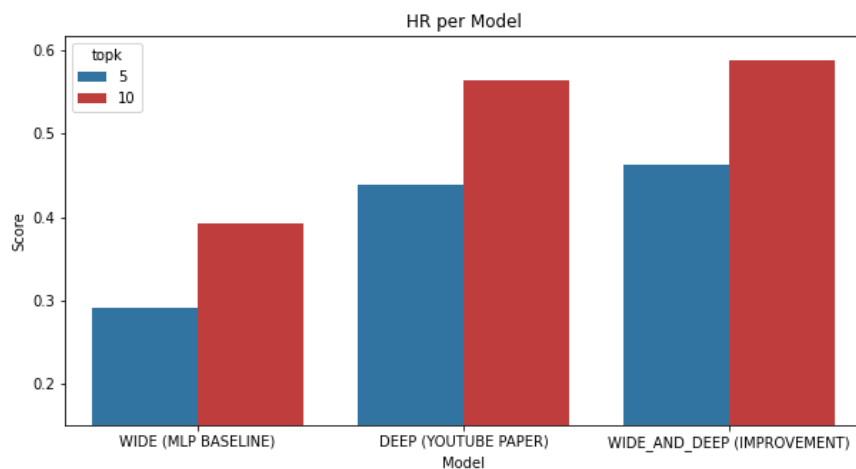
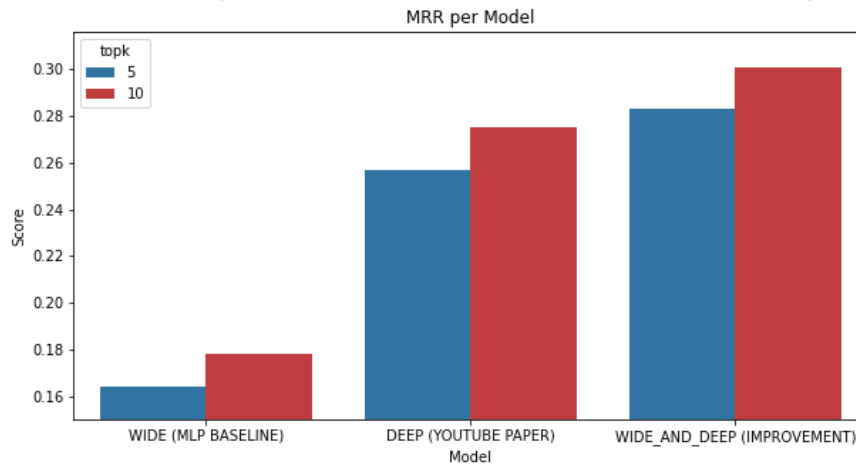
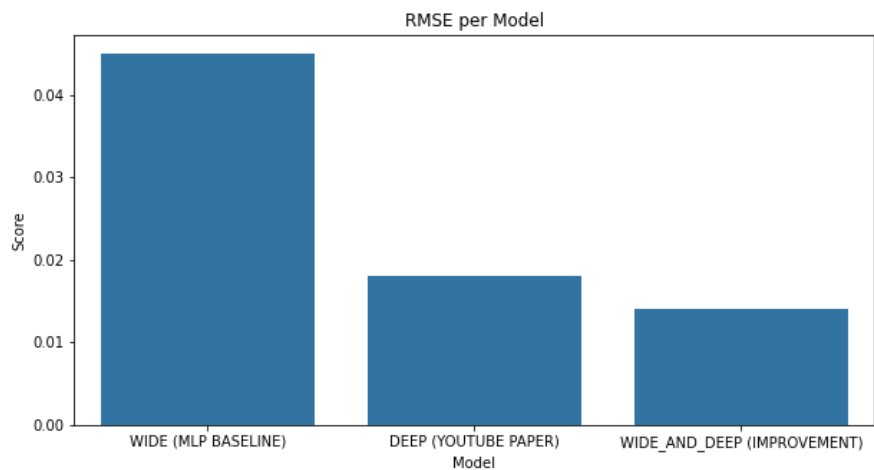


## Results:

For the candidate phase the output is 10 candidate movies for each user, thus, we've used HR@5, HR@10, MRR@5, MRR@10, NDCG@5, NDCG@10. We can see that the WIDE and DEEP model outperformed the DEEP model, which in turn outperformed the WIDE model.



For the ranking phase, the output is a scalar for the rating score, thus, we've used RMSE. As in the candidate phase, we can see that the WIDE and DEEP model outperformed the DEEP model, which in turn outperformed the WIDE model.



The full comparison table is as follows:

	model	topk	metric	score
0	WIDE (MLP BASELINE)	5	HR	0.291
1	WIDE (MLP BASELINE)	10	HR	0.393
2	WIDE (MLP BASELINE)	5	MRR	0.164
3	WIDE (MLP BASELINE)	10	MRR	0.178
4	WIDE (MLP BASELINE)	5	NDCG	0.195
5	WIDE (MLP BASELINE)	10	NDCG	0.229
6	DEEP (YOUTUBE PAPER)	5	HR	0.438
7	DEEP (YOUTUBE PAPER)	10	HR	0.565
8	DEEP (YOUTUBE PAPER)	5	MRR	0.257
9	DEEP (YOUTUBE PAPER)	10	MRR	0.275
10	DEEP (YOUTUBE PAPER)	5	NDCG	0.302
11	DEEP (YOUTUBE PAPER)	10	NDCG	0.344
12	WIDE_AND_DEEP (IMPROVEMENT)	5	HR	0.462
13	WIDE_AND_DEEP (IMPROVEMENT)	10	HR	0.588
14	WIDE_AND_DEEP (IMPROVEMENT)	5	MRR	0.283
15	WIDE_AND_DEEP (IMPROVEMENT)	10	MRR	0.301
16	WIDE_AND_DEEP (IMPROVEMENT)	5	NDCG	0.328
17	WIDE_AND_DEEP (IMPROVEMENT)	10	NDCG	0.369

	model	metric	score
0	WIDE (MLP BASELINE)	RMSE	0.045
1	DEEP (YOUTUBE PAPER)	RMSE	0.018
2	WIDE_AND_DEEP (IMPROVEMENT)	RMSE	0.014