A Recommendation System to Enhance Midwives' Capacities in Low-Income Countries

Anna Guitart, Afsaned Heydari, Eniola Olaleye, Jelena Ljubicic,
Ana Fernández del Río, África Periáñez and Lauren Bellhouse
benshi.ai, Barcelona, Spain, Maternity Foundation, Copenhagen, Denmark

benshi.ai MATERNITY foundation BILL & MELINDA GATES foundation

Abstract

Maternal and child mortality is a public health problem that disproportionately affects low- and middle-income countries. Every day, 800 women and 6,700 newborns die from complications related to pregnancy or childbirth. And for every maternal death, about 20 women suffer serious birth injuries. However, nearly all of these deaths and negative health outcomes are preventable. Midwives are key to reverting this situation, and thus it is essential to strengthen their capacities and the quality of their education.

This is the aim of the **Safe Delivery App, a digital job aid and learning tool** to enhance the knowledge, confidence, and skills of health practitioners. Here, we use the behavioral logs of the App to implement a recommendation system that presents each midwife with suitable content to continue gaining expertise. We focus on predicting the click-through rate, the probability that a given user will click on a recommended content. We evaluate four deep learning models and show that all of them produce highly accurate predictions.

Introduction

Most maternal and newborn deaths, and other negative health outcomes, occur in low- and middle-income countries and are preventable through timely management by a skilled health professional. Access to quality midwifery would improve the long-term health and welfare of mothers and babies [1].

Recommendation systems have been used in many contexts, including healthcare applications [2], where the most typical approaches are collaborative filtering or content-based filtering. Here we explore such a recommendation system, focusing instead on the **click-through rate (CTR) prediction**, namely on estimating each user's probability of clicking on a certain App content. Once identified, the content that each user is more likely to check in the near future can be made more accessible. This should improve the user experience and accelerate the acquisition of critical skills.

Method

The dataset was extracted from the **Maternity Foundation's Safe Delivery App** and consisted of user **behavioral logs** collected between 2019-01-01 and 2021-09-01 (which included 21296 users from **India** and 1486 from **Ethiopia)**. The App accommodates different learning sections grouped in modules. We focused on videos organized in different chapters and the drugs section.

Our strategy involves predictive modeling to foresee a user's response to potential recommendations. For a certain content, the CTR is the rate at which a user presented with the opportunity to consume it will do so. We focus on predicting if a user will check some specific content the next day, a problem to which we can apply advanced CTR prediction models. The main challenges of these models are dealing with high-dimensional, very sparse feature spaces and capturing feature interactions.

We employed four different models and compared their performance: **Product-based neural networks** (PNNs) [3], **Deep factorization machines** (DeepFM) [4], **Extreme deep factorization machines** (xDeepFM) [5] and **Dual input-aware factorization machines** (DIFM) [6].

Results

The models predict the probability that a given user will click on a specific content (either a individual drug, a drug family, a module video or a specific video chapter). To evaluate model performance, we treated predictions as a binary classification and calculated area under the curve (AUC) [7]. Additionally, we calculated the root-mean-square error (RMSE). Results are shown in Table 1 (AUC) and Table 2 (RMSE).

All models show good performance for all types of content (AUC above 0.9 in almost all instances and low RMSE). Although models show better performance for the drug families than for individual drugs, the difference is not significant. This means that our models are able to foresee not only the health problems that will be encountered, but also the specific drug that will be checked. Even with the smaller training-sample sizes in the Ethiopian case our models still managed to learn and give accurate predictions.

The fact that the xDeepFM and DIFM models did not perform significantly better than the simpler PNN and DeepFM ones suggests that **high-order vector-wise feature interactions are not relevant in this problem** for the selected features. The DIFM model produces the best AUC values for drugs, while consistently yielding the worst results for videos. This could mean that **input-aware reweighting could have some importance in the drug case.**

RMSE in Table 2 shows that the DIFM model is always among the worst choices, while **DeepFM and xDeepFM errors are very similar and the lowest ones for all content types.** This means DeepFM and xDeepFM should outperform the rest by tuning the probability threshold above which a positive occurrence is predicted.

This is further supported by Figures 1 and 2, which show the RMSE of the individual-drug predictions produced by each model for Ethiopia and India. These figures reveal that the error is not uniformly distributed, with the largest contributions coming from only a few drugs. They show, too, that **there is a group of drugs for which all models present large errors, with the PNN and DIFM models exhibiting additional peaks.**

Table 1.
Area under the ROC curve (AUC) for each model (PNN, DeepFM, xDeepFM and DIFM) and each of the recommender targets (drug, drug family, video chapter and video module). The values were calculated using datasets from Ethiopia and India.

		Model			
Dataset	Country	PNN	DeepFM	xDeepFM	DIFM
Drug	Ethiopia	0.9519	0.9684	0.9674	0.9766
	India	0.9694	0.9739	0.9587	0.9735
Drug Family	Ethiopia	0.9853	0.9310	0.9696	0.9700
	India	0.9664	0.9717	0.9645	0.9742
Video Chapter	Ethiopia	0.8727	0.8814	0.9070	0.8199
	India	0.9654	0.9712	0.9692	0.9363
Video Module	Ethiopia	0.9074	0.9019	0.9055	0.8653
	India	0.9657	0.9752	0.9727	0.9319

Table 2.
Root-mean-square error (RMSE) for each model (PNN, DeepFM, xDeepFM and DIFM) and each of the recommender targets (drug, drug family, video chapter and video module). The values were calculated using datasets from Ethiopia and India.

		Model			
Dataset	Country	PNN	DeepFM	xDeepFM	DIFM
Drug	Ethiopia	0.0886	0.0824	0.0832	0.1089
	India	0.0330	0.0263	0.0256	0.0404
Drug Family	Ethiopia	0.1219	0.1216	0.1205	0.1362
	India	0.0487	0.0376	0.0373	0.0586
Video Chapter	Ethiopia	0.1587	0.1029	0.1042	0.1289
	India	0.1375	0.0555	0.0553	0.1113
Video Module	Ethiopia	0.1857	0.1789	0.1789	0.2504
	India	0.1055	0.0929	0.0939	0.1955

Conclusions

Results have shown that **state-of-the-art CTR prediction models can successfully foretell whether a user will check a certain video or drug content within the next day.** Therefore, they can be used to make that content more accessible to the user. The models could be used to facilitate midwives' everyday practice and skill acquisition, ultimately resulting in a better care for the mothers and babies they are assisting.

Results suggest that the simpler DeepFM architecture would suffice to address this content recommendation problem in a production environment. This seems to indicate that explicitly considering high-order vector-wise feature interactions and feature-aware reweighting has no significant effect in this particular problem. On the whole, in terms of overall performance and efficiency, the **DeepFM model appears like the ideal candidate to be used in production.**

Contact: Olaleye Eniola / eniola@benshi.ai / www.benshi.ai

Figure 1. Root-mean-square error (RMSE) per drug for the predictions of the PNN, DeepFM, xDeepFM and DIFM models across all users in Ethiopia.

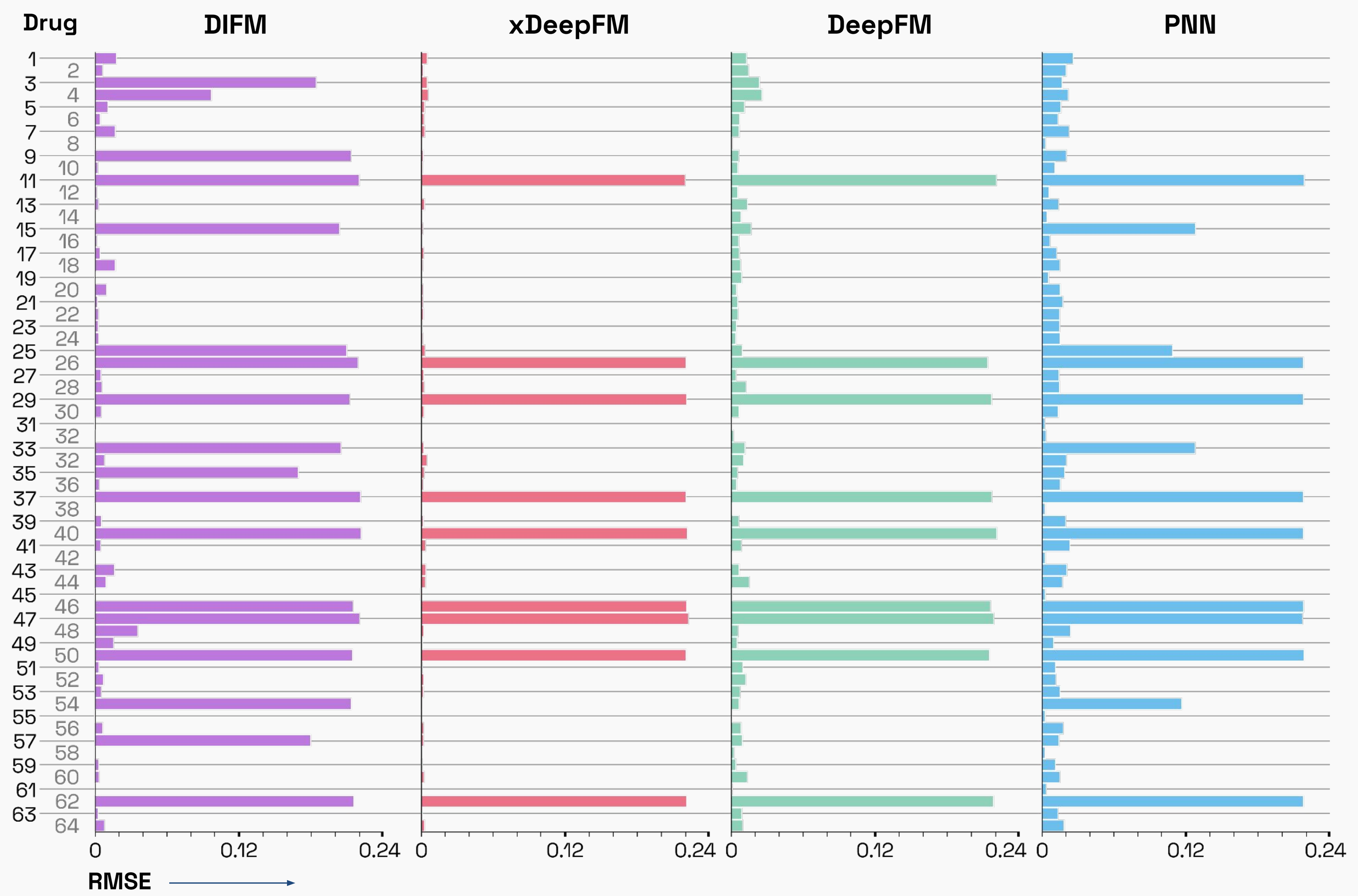
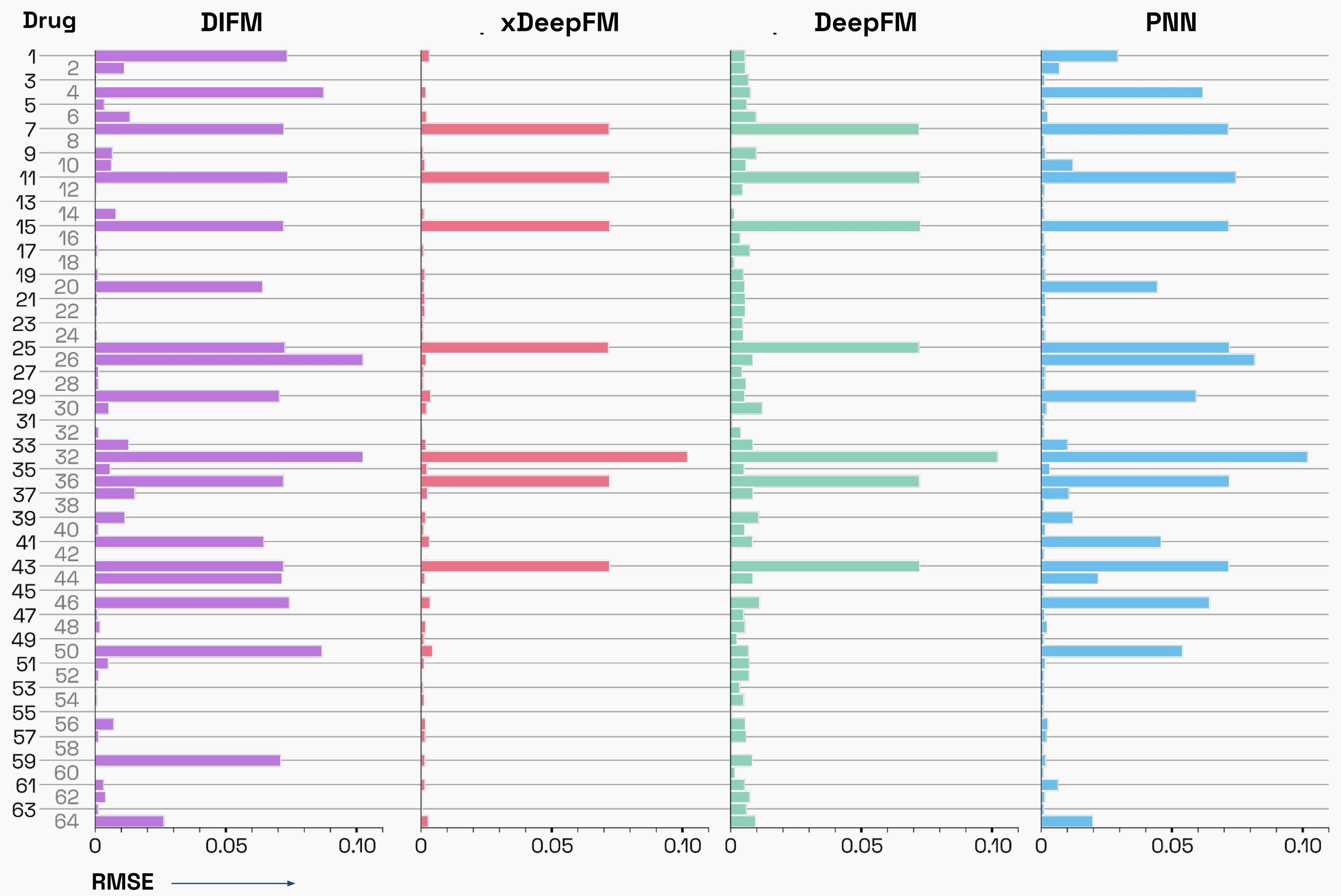


Figure 2. Root-mean-square error (RMSE) per drug for the predictions of the PNN, DeepFM, xDeepFM and DIFM models across all users in India.



Drug Key for Figure 1 & Figure 2

1. adrenaline
2. aminophylline
3. amoxicillin
4. ampicillin
5. antibiotic-eye-drops
6. artemether-20-mg-lumefantrin-120-mg
7. artesunate
8. atropine
9. betamethasone
10. caffeine-citrate
11. calcium-gluconate
12. carbetocin
13. cefazolin
14. cefotaxime
15. ceftriaxone
16. cefuroxime
17. chloramphenicol
18. chlorhexidine
19. cloxacillin
20. contraceptives-birth-control-pills
21. contraceptives-implant
22. contraceptives-injectable-depo-provera
23. contraceptives-intrauterine-device
24. contraceptives-overview
25. dexamethasone
26. diazepam
27. diclofenac
28. dicloxacillin
29. ergometrine
30. erythromycin
31. fentanyl
32. flucloxacillin
33. gentamicin
34. hydralazine
35. ibuprofen
36. indomethacin
37. iron
38. ketamine
39. labetalol
40. lidocaine
41. magnesium-sulphate
42. mefloquine
43. methyldopa
44. metronidazole
45. mifepristone
46. misoprostol
47. morphine
48. nifedipine
49. nitrofurantoin
50. oxytocin
51. paracetamol
52. penicillin-g-benzylpenicillin
53. pethidine
54. phenobarbital
55. piperaquine
56. post-exposure-prophylaxis---pep
57. quinine
58. sulfamethoxazole-with-trimethoprim
59. syntometrine
60. tetracycline
61. tramadol
62. tranexamic-acid
63. trimethoprim
64. vitamin-k

References

1. Andrea Nove, Ingrid K Friberg, Luc de Bernis, Fran McConville, Allisyn C Moran, Maria Najjemba, Petra ten Hoope-Bender, Sally Tracy, and Caroline SE Homer. *Potential impact of midwives in preventing and reducing maternal and neonatal mortality and stillbirths: a lives saved tool modelling study. The Lancet Global Health, 9(1):24–32, 2021*
2. *Luciano Rodrigo Ferretto, Cristiano Roberto Cervi, and Ana Carolina Bertoletti de Marchi. Recommender systems in mobile apps for health a systematic review. In 2017 12th Iberian Conference on Information Systems and Technologies (CISTI), pages 1–6. IEEE, 2017.*
3. *Yanru Qu, Han Cai, Kan Ren, Weinan Zhang, Yong Yu, Ying Wen, and Jun Wang. Productbased neural networks for user response prediction. In 2016 IEEE 16th International Conference on Data Mining (ICDM), pages 1149–1154. IEEE, 2016.*
4. *Huifeng Guo, Ruiming Tang, Yunming Ye, Zhenguo Li, and Xiuqiang He. Deepfm: a factorization-machine based neural network for ctr prediction, 2017.*
5. *Jianxun Lian, Xiaohuan Zhou, Fuzheng Zhang, Zhongxia Chen, Xing Xie, and Guangzhong Sun. xdeepfm: Combining explicit and implicit feature interactions for recommender systems. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1754–1763, 2018*
6. *Wantong Lu, Yantao Yu, Yongzhe Chang, Zhen Wang, Chenhui Li, and Bo Yuan. A dual input-aware factorization machine for ctr prediction. In IJCAI, pages 3139–3145, 2020.*
7. *James A Hanley and Barbara J McNeil. The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology, 143(1):29–36, 1982.*