

# SMART TUNES FOR KIDS

## Comparing Deep Learning with Traditional Models in Music Recommendations for Children



### INTRODUCTION

When you access almost any online platform, is content you might be interested in suggested to you. This is the work of recommender systems (RS). These RS are often designed with just adults in mind and their main objective is to keep you interacting with the platform. Children however also use these online platforms on a daily basis but have other needs than just watching content that keeps them engaged. These are needs such as more educational content or a diverse and novel set of recommendations to stimulate their curiosity. We do not know whether deep learning models are capable of fulfilling such needs and explore their capabilities in this study.

### RESEARCH QUESTION

Do music recommendation systems using deep learning models outperform traditional models in terms of performance, diversity, and novelty when trained and evaluated on child data?



### METHODOLOGY

We use the list of songs underage users of the Last.FM music streaming platform have interacted with. We then split the data with into a test set and validation set. The models are trained with the test set and their recommendations are compared against the validation set. Consequently are the values of each metric computed, to provide us with an insight of how the models score on different aspects.

Performance, diversity and novelty are each aspects a RS that is suitable for children requires. Performance metrics measure how many songs the model recommends that are in the validation set, diversity measures how diverse the recommendations are and novelty measures how popular the recommended songs are.

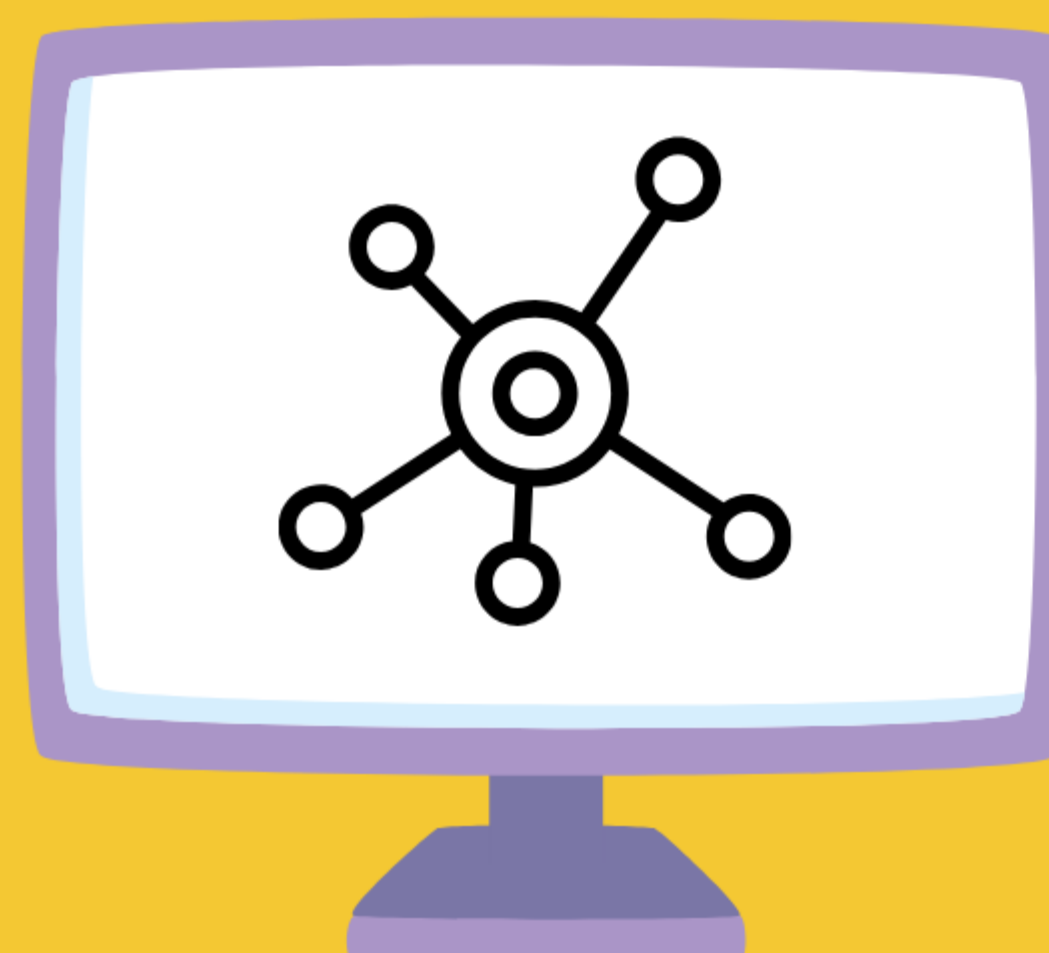
#### Metrics:

##### Performance

- nDCG → Overall accuracy
- MRR → First accurate item

##### Beyond-performance

- Gini Index → Diversity
- EPC → Novelty



Our selection for baseline models consists of basic benchmarks and models that are known to be well performing and utilize a wide range of traditional recommendation techniques.

#### Baseline (traditional) models:

- Random
- MostPop
- BPRMF
- UserKNN

Our selection for deep learning models consists of state-of-the-art models that each utilize a different form of deep learning, therefore giving us insight into the capabilities of different aspects of deep learning

#### Deep learning models:

- NeuMF
- MultiVAE
- AMF
- NGCF

We did not employ hyperparameter optimization for this study, as to observe the baseline performance of the models and conduct a fair comparison under standardised conditions. We did utilize basic hyperparameter optimization for our best performing deep learning model MultiVAE, shown in the results as MVAE HP, as to give a general idea of the capabilities of deep learning when slightly optimized.

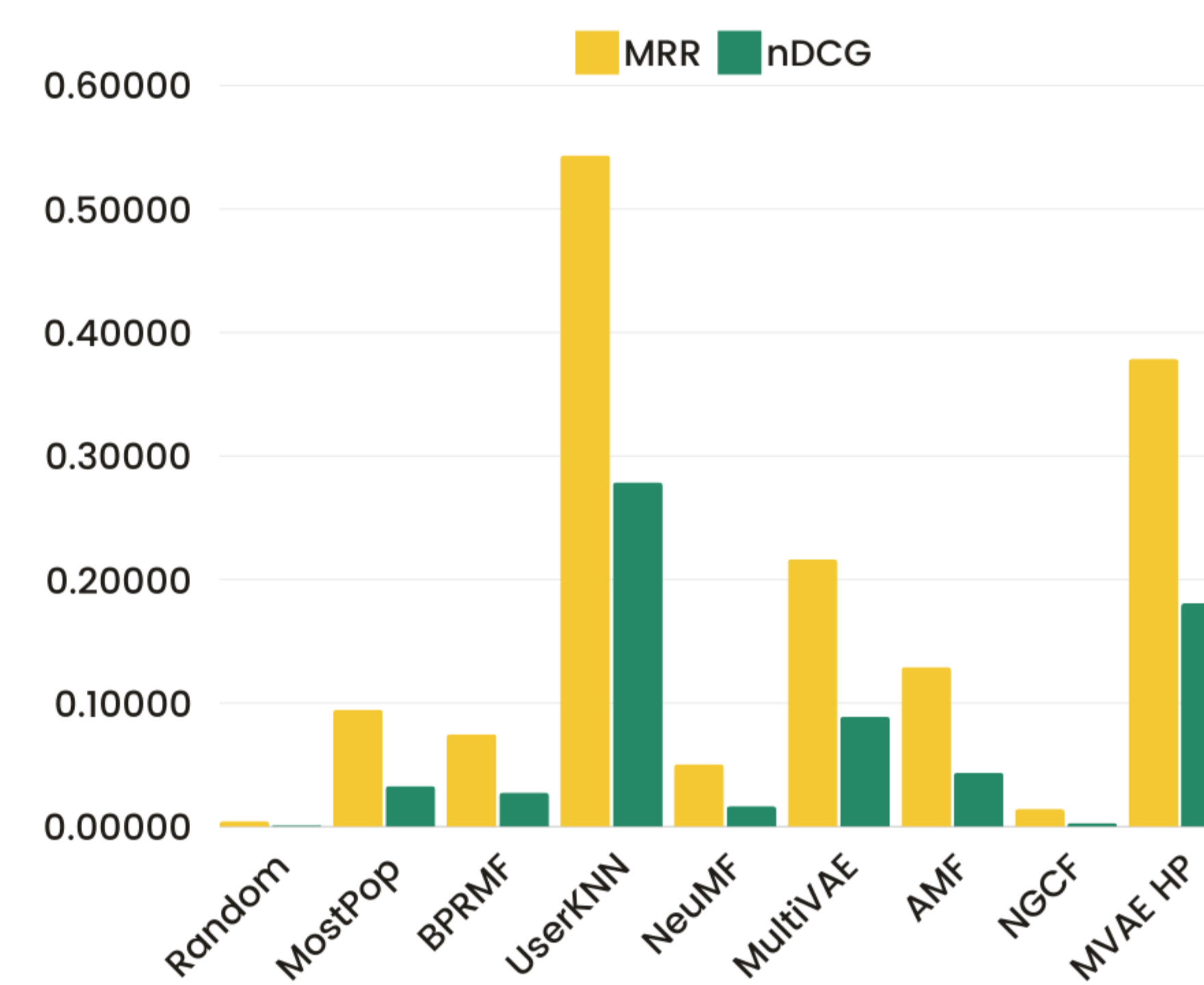
### AUTHOR

Lennart Verstegen  
l.h.verstegen@student.tudelft.nl

Supervisors: Dr. Maria Soledad Pera, Robin Ungruh



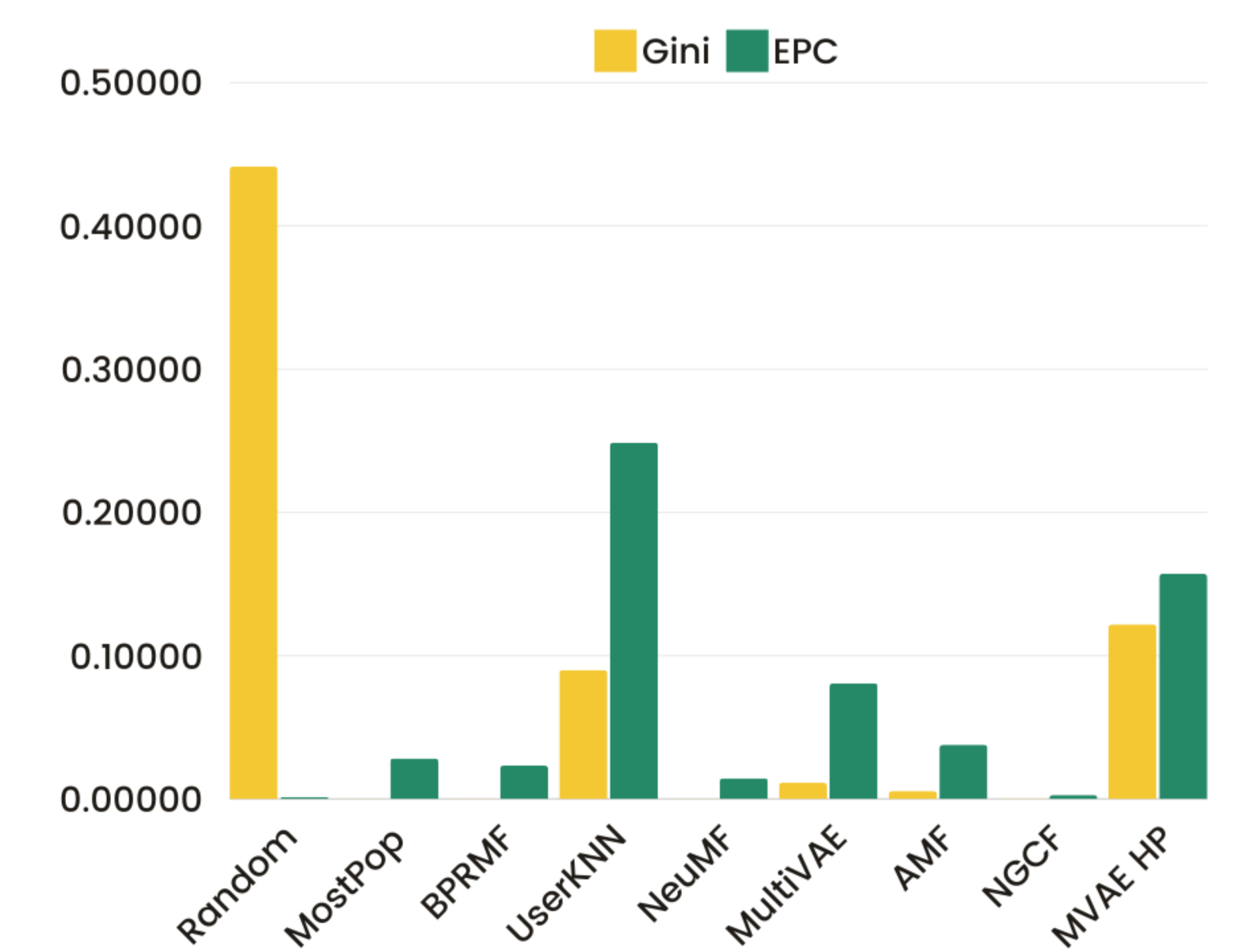
### RESULTS



Performance metric results

Most models performed quite poorly. This is due to all of the models not utilizing hyperparameter optimization. UserKNN is the clear winner of the performance metrics, with MVAE HP as strong second.

Noticable is how UserKNN and MVAE HP are the highest scoring models in both performance and beyond-performance models.



Beyond-performance metric results

In terms of the beyond-performance metrics, did MVAE HP have the best diversity and UserKNN the best novelty. We do not count the high diversity value of Random.

### CONCLUSION

#### Key findings:

- Deep learning models do not outperform established traditional models
- The baseline performance of deep learning models is very poor for recommending to children
- Nearest neighbourhood models excel in the context of music recommendation for children
- There is large potential in deep learning models when hyperparameter optimization is employed

### FUTURE WORK

There is a plethora of ways this study can be expanded upon. As there is still potential to be seen in deep learning models, can future studies employ hyperparameter optimization as to explore their full capabilities. As we only covered a small set of characteristics of a RS suited towards children, should these models be compared against more metrics encompassing those characteristics. This study only employed the listening history of users, but future studies can explore the performance of a multitude of models that use other features such as song features.

