# Evaluating the Performance of Different Models for Children's Book Recommendations

**Ziang Qiu**[1]

**Supervisor: Sole Pera**[1]

[1]EEMCS, Delft University of Technology, The Netherlands

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
January 29, 2023

Name of the student: Ziang Qiu
Final project course: CSE3000 Research Project
Thesis committee: Sole Pera,Willem-Paul Brinkman

## Abstract

Recently, a few children-centered recommendation systems have been created and evaluated. However, these systems required user interaction to create ground truth to evaluate the result. This research aims to compare some of the traditional recommendation models and explore which trait could impact the recommendation process most for different age group users. The result shows the children friendly model does not achieve higher accuracy than the traditional recommendation model. But the book length model and emotion analysis model shows the potential of a good RS that can help children choose books, and cover image recommendation models are only working with younger age.

## 1 Introduction

Recommendation systems (**RS**) have become increasingly popular during past decades. It generates valuable user recommendations based on multiple preferences, and models [1]. There are two main categories of RS: content-based RS [2] and collaborative filtering RS [3]. These RS wildly used by different websites and companies to recommend interesting products to different groups of users [4, 5]. It is also operated in multiple domains to raise company profit or attract users' interest, such as movies and grocery items [4, 6]. Book is also one of the most popular domains that RS is working with. Wang et al. [7] suggest that book RS could help people with many aspects, such as retrieving required documentation and suggesting exciting books. While most RS is designed for adults, there are limited numbers of RS designed for children. It may cause some children not to find an interesting book on the bookshop shelf. More and more children give up on reading books and spending time on mobile entertainment[8][9].

Hosokawa et al.[10] mention that children spend more time on mobile devices than they ever used to. It has been noticed that children's reading ability has dropped since the pandemic started [11, 12]. Mondal et al. [13] find addiction to mobile devices and the internet could harm academic performance. Hu et al. [14] find out that reading improves kids' vocabulary skills and concentration levels. However, some limited websites or systems can make suitable recommendations to children based on their preferences [15]. Many RS suggest books based on users' social tags or browser history. Because of privacy issues and limited survey abilities [16], it is almost impossible to collect underage users' data for generating a recommendation list. Most of these models are designed for adults instead of children. However, a children-centered RS could be implemented with some traits of the book. Milton et al. [17] suggest that it is possible to recommend books based on the properties, such as a cover image or introduction text. It is still being determined if other traits could influence the performance of the book RS for children.

This research is designed to compare some traditional recommendation algorithms when targeting different age groups of users and explore which trait could impact the recommendation process most in RS. In this research, the target group users mainly are children. The age groups are divided by education level, including kindergarten, elementary, and middle school. The adult group has been created as the control group. The current researches generate the read-like list by human interaction. None of them are tested with real-world reader-like data. This research implements these algorithms and tests them with a real-world dataset with MRR and Hits@5 metrics. The remainder of this paper is structured as follows. The second section reviews current existing work. Section 3 will describe the method and dataset used in this research. After that, Section 4 will discuss the experiment setup, and result, and give a summary and brief discussion of the result. In the last, Section 5 will provide a conclusion and discuss the future work of this research.

## 2 Background and Related Work

In this section, we discuss related literate that informs our work.

### 2.1 Recommendation Techniques

RS is a subsystem of the information retrieval system [18] to predict the users' favorites on a specific item. RS is deployed in various areas, such as product recommenders for online stores [4] or entertainment content recommenders for social media platforms [19]. There are several different types of recommendation techniques. Content-based filtering [2]. This model generates recommendations based on the traits of the item being recommended. Collaborative filtering [3, 20] makes recommendations based on the historical behavior of similar users, such as ratings or reviews on the same item. A hybrid approach [21] mixes both the content-based filtering models and the collaborative filtering models to generate recommendations. This research used both content-based filtering models and collaborative filtering models to generate a recommendation list.

### 2.2 Recommendation Evaluation

The performance of a recommendation algorithm can be evaluated using various metrics. The metrics utilized depend on the filtering method employed in the RS[22]. *Accuracy* calculates the rate of recommended items is accepted value [23], while *coverage* evaluates the percentage of items that can be generated recommendation by the RS[24]. Both of these two measurements are useful for assessing the quality of a recommendation algorithm, as they can help to determine how well the algorithm can meet the needs and preferences of users.

Statistical accuracy metrics and decision support accuracy metrics are the two main classes to evaluate the accuracy of the RS [20].

#### Statistical Accuracy Metrics

*Statistical Accuracy Metrics* evaluates the difference between the predicted result value and the true value[25, 26]. Root Mean Square Error (**RMSE**) and Mean Average Error(**MAE**) are the most popular metrics among statistical accuracy metrics. MAE measures the average value of absolute error between the expected value and prediction score. It calculates as Equation 1:

$$MAE = \frac{1}{|D|}\sum_{i=1}^{|D|}|x_i - y_i| \quad (1)$$

Where $x_i$ is the predicted value generated by the RS for item $i$, and $y_i$ is the expected value for item $i$ in the dataset. $|D|$ is the size of the dataset. A higher MAE means worse accuracy for the RS generated recommendations.

**Decision Support Accuracy Metrics**

Precision, Recall, False Positive Rate, Specificity, and Mean Reciprocal Rank (**MRR**) are the most common metrics in *Decision Support Accuracy Metrics*. It generates the effectiveness of the RS in choosing items from the set of candidates. These metrics are based on the assumption that the prediction process is a binary operation. The outcome after the evaluation could only be 4 different results in Table 1 [25].

| | Recommended | Not Recommended |
|---|---|---|
| Related | True-Positive (TP) | False-Negative (FN) |
| Unrelated | False-Positive (FP) | True-Negative (TN) |

Table 1: Acceptable Outcomes for Decision Support Accuracy Metrics

Precision evaluates the percentage of recommended items that are related to the user, and recall measures the capability of a model or system to identify all of the relevant items in a dataset. They are calculated in the following formula:

$$Precision = \frac{|TP|}{|TP| + |FP|} \quad (2)$$

$$Recall = \frac{|TP|}{|TP| + |FN|} \quad (3)$$

False Positive Rate calculates the rate at a negative sample categorized as positive, and Specificity shows the system's ability to find negative samples. The formulas for False Positive Rate and Specificity are:

$$False\_Positive\_Rate = \frac{|FP|}{|FP| + |TN|} \quad (4)$$

$$Specificity = 1 - False\_Positive\_Rate \quad (5)$$

MRR calculates the reverse of the first relative item position in the recommendation list[27]. The formula of MRR looks like Equation 6:

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i} \quad (6)$$

Where $rank_i$ is the rank of the first relevant item in the recommendation list and $|Q|$ stands for the length of the dataset. The higher MRR suggests a better recommendation list generated by the RS.

These two types of metrics could help this research to analyze the recommendation list.

### 2.3 Children's book recommendation system

Some children-centered book recommendation systems have been described in the literature in the past [15, 28–30]. CBRec [15] suggests a hybrid approach with a combination of different filtering models and shows an improvement in making book recommendations to children. It compares various existing models and their combinations. StoryTime [29] creates a book recommendation model which does not rely on the user's historical data or adult guidance. With user input, it automatically generates recommendations and displays them with a user-friendly GUI. Rabbit [30] generates personal book recommendations to children based on their literacy level and personal interests, including topic and content. However, it is not possible to compare their performance with each other. All of these RS required user-system interaction that can serve as ground truth [16]. It is not possible to compare them with real-world user-like data.

At the same time, recent research from Milton et al [17] has identified few book traits that could potentially be used as indicators to recommend books for children. This research has found younger children prefer brighter cover images, positive emotions, and books with fewer pages. These findings are never verified with real-world data and it may possible to build a children friendly RS with these traits. This research intends to test the children friendly recommendation models with widely used RS, including collaborative filtering model based on popularity and content-based filtering model based on text similarity.

## 3 Method

The target of this research is to find the performance of existing recommendation algorithms when targeting non-mainstream user groups and explore which trait could impact the recommendation process most in RS. To reach this target, this research used the GoodReads dataset to create and assess different recommendation algorithms with two evaluation metrics.

### 3.1 Algorithms

This research employed several widely-used recommendation models and some children-centered recommendation algorithms.

- *Popular* recommending books based on popularity, including text review counts and rating counts. =
- *Description Similarity*, an item-based collaborative filter using term frequency–inverse document frequency function[31] and cosine similarity function to calculate the similarity score. Cosine similarity between two vectors is calculated by equation 7.

$$\cos(\theta) = \frac{A \cdot B}{\|A\|_2 \|B\|_2} \quad (7)$$

- *Emotions* an item-based collaborative filter using emotions similarity purposed by Marko et al.[32]. In this research, the Text2Emotion [1] package is used to analyze

[1]https://pypi.org/project/text2emotion/

the book's description. It generates a vector representing the emotion in multiple degrees. The cosine similarity function measures the similarity of the two books vectors.

- *Book Length* an item-based collaborative filter using the book-length similarity to generate recommendations[33].

- *Book Covers* an item-based collaborative filter using the cover image property to generate recommendations. Brightness and colorfulness are used in this research to compute the similarity. Details on calculating these properties can be discovered in [17]. A book with the most colorful and bright cover image will be recommended.

This research does not compare the performance between different recommendation models but finds the performance of existing recommendation algorithms when targeting non-mainstream user groups and explore which trait could impact the recommendation process most in RS.

### 3.2 Data Set

Due to the privacy laws[34] for underage, there are not many data sources available for use in the research. In late 2017, Wan et al. [35, 36] created the GoodReads dataset from the website *GoodReads* [2] with users' public shelves. It comes up with metadata and user interaction such as average ratings, review counts, and rating counts. The GoodReads data source contains 2.3M records of books and 228M user-book interactions from 542K different users. Each instance has an ISBN, author, review counts, popular shelves, average rating, similar books, introduction, and cover image.

This research generates a new data GoodReadsSample (**GDS**) set based on the GoodReads data source. This research analyzed the results with Decision Support Accuracy Metrics, and the ground truth is needed. The user generates the similar book property in the data source and contains the most user-voted similar books in the list. The similar book list will be used as ground truth and reader-like data in this research, and the instances in the data source without a similar book list will be removed. This research needs to use cover images to generate recommendations. Any book without a cover image will be removed from the dataset and data source. The popular shelves field in the data source indicated the target group of readers. With this property, the dataset will be divided into four sub-datasets by age[37], including kindergarten (¡=5), elementary (6-10), and middle school (11-13). Each instance in these datasets should only contain only one tag related to age in the popular shelves list to prevent overlap. As a control group, this research generates another sub-dataset for adults. The graph 1 and table describe the dataset.

[2] https://www.goodreads.com/

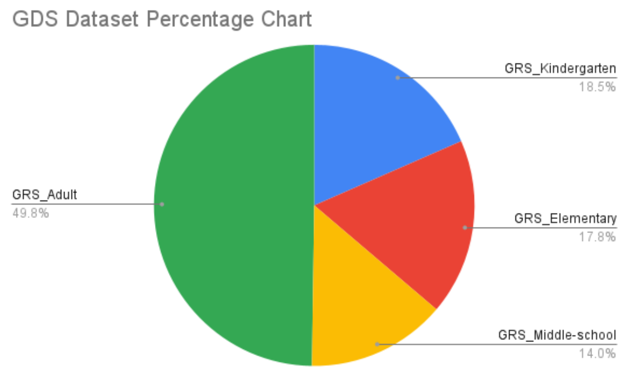


Figure 1: GDS Dataset Percentage Chart

| Name | #Book | Average Similar Books |
|---|---|---|
| GDS | 20092 | 8.11 |
| GDS_Kindergarten | 3708 | 7.84 |
| GDS_Elementary | 3567 | 6.85 |
| GDS_Middle_School | 2817 | 10.68 |
| GDS_Adult | 10000 | 7.94 |

Table 2: GDS statics table

The overall dataset contains 20k books with an average of 8.11 similar books. The kindergarten dataset has 3.7k books and an average of 8.44 similar books. The elementary and middle school dataset contains 3.5 k and 2.8 k books, with 6.73 and 5.91 similar average books. In comparison, this research generates an adult dataset including 10k books with an average of 8.4 similar book lists. Graph 2 shows the distribution of similar books for these datasets.

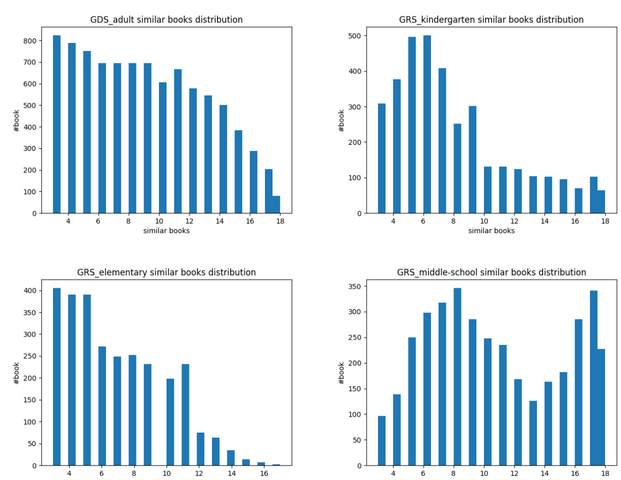


Figure 2: Similar book distribution

### 3.3 Candidates Generation

The candidate list is a list containing books for RS input. In this research, candidate generation combines the ground truth with the books in the data source. The average ground truth list for GDS is 8.11. For each instance in the dataset, RS will generate 100 candidates. To prevent bias in the book with more ground truth. Any instance in the dataset with more than 8 books in a similar list will be randomly filtered until 8 books are left over. After filtering the similar list, the RS will randomly select 92 books from the instance's category and add them to the candidate list.

### 3.4 Performance Metrics

To measure the performance of the recommendation system, two wildly accepted metrics are implemented in this research, including MRR and hits@k. MRR has been introduced in Section 2.2. Hits@k is another widely used decision support accuracy metric. Hits@k calculated the score based on the top-K recommendations items found in the ground truth list. The formula for Hits@k is:

$$Hits@K = \frac{1}{|D|}\sum_{i=1}^{|D|} H(D_i, k) \tag{8}$$

Where $D$ is the recommendation list RS generated, $H(D_i, k)$ returns 1 if one item in the top-K recommendation list has been founded in the ground truth list. Otherwise, it returns 0. A higher Hits@k score shows the better quality of the recommendation list.

## 4 Experiment Result and Discussion

This research implements the data, recommendation algorithms, and metrics presented in Section 3. This section's target is designing to answer the following research questions:

RQ1 Does the same RS perform differently if it targets different age groups?

RQ2 What is the traditional model's performance compared to children-friendly models when applying children's books?

RQ3 which trait could impact the recommendation process most in RS?

### 4.1 Experiment Results

In order to evaluate the performance of existing recommendation algorithms when targeting children, this research conducted an experiment that compares the performance of traditional recommendation algorithms with children-friendly algorithms for different age groups. Two metrics in section 3.4 analyze the results. Table 3 shows the recommendation result obtained by Hits@5 metrics, whereas Figure 3 and Figure 4 show the MRR result for all recommendation models.

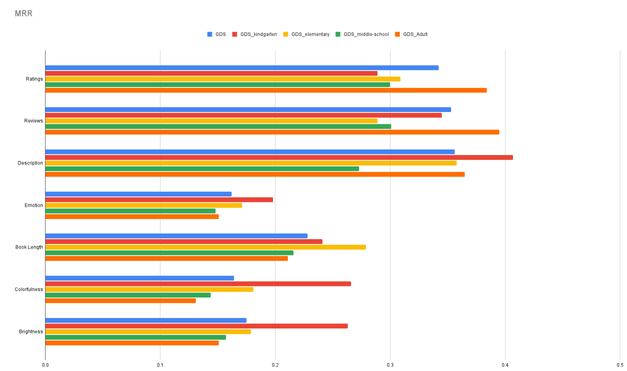


Figure 4: MRR computed for different recommendation models across datasets capturing users of different ages

Experiment results show that the RS performs differently in age groups, especially in kindergarten.

**Popular**

From Figure 4, the average score of the adult dataset is higher than others in both the ratings model and reviews model. It suggests the popular model works better for the adult group. The leftover three sub-datasets got almost the same result in this metric. However, Table 3 shows an opposite signal. Comparing the results for the different sub-datasets, it appears that the kindergarten sub-dataset has the highest value in both the ratings model and reviews model, with 0.592 and 0.632, respectively.

Based on the values in the two metrics result, the overall performance of the system evaluated on the middle school, and elementary sub-dataset is relatively lower than other sub-datasets in both the "Ratings" and "Reviews" columns. It is harder to generate a correct recommendation for the children based on popularity

It is also worth noting that the difference between the values of the rating model and review model for the overall dataset is not huge, which means the system performs stably between the two properties.

**Description**

The description model accuracy is generally slightly lower than the model based on popularity. It is more evident in the middle school dataset. The elementary and adult group has almost the same performance on both metrics. It suggests description model performs similarly in these two groups. MRR metric shows the kindergarten group achieves the highest score among all models, and the hits@5 metric shows the best accuracy among all datasets in this model. It suggests that elementary children prefer books with similar introductions, which is less effective when they grow up.

**Emotion**

For the emotional similarity-based model in the book description, the Hits@5 table shows GDS_kindergarten $>$ GDS_middle-school $>$ GDS_adult $>$ GDS_elementary. However, in the MRR table, the GDS_elementary is not the worst. It ranked second in the MRR result table. It suggests that

| Dataset Name | Ratings | Reviews | Description | Emotion | Book Length | Colorfulness | Brightness |
|---|---|---|---|---|---|---|---|
| GDS | 0.543 | 0.575 | 0.517 | 0.335 | 0.367 | 0.242 | 0.272 |
| GDS_kindgarten | 0.592 | 0.632 | 0.58 | 0.364 | 0.398 | 0.33 | 0.418 |
| GDS_elementary | 0.454 | 0.49 | 0.513 | 0.308 | 0.475 | 0.277 | 0.270 |
| GDS_middle-school | 0.475 | 0.545 | 0.409 | 0.326 | 0.339 | 0.239 | 0.264 |
| GDS_adult | 0.579 | 0.595 | 0.533 | 0.338 | 0.33 | 0.201 | 0.208 |

Table 3: Experiment Result with Hits@5 Metrics

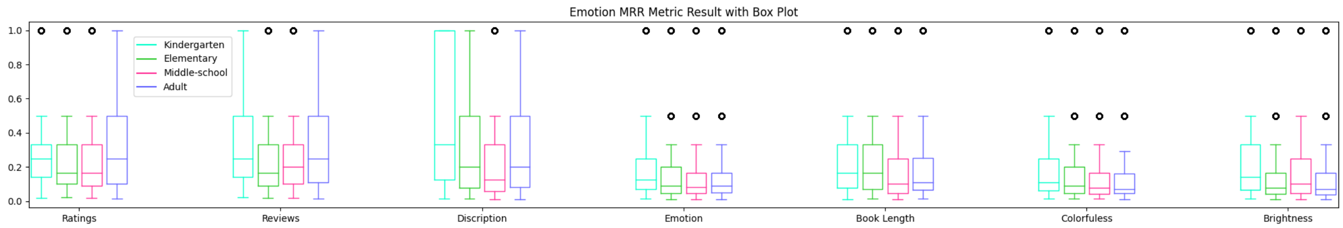


Figure 3: MRR Metric Result with Box Plot

the elementary group performs poorly in the top 5 recommendation list. However, the elementary group's overall performance is better than GDS_middle-school and GDS_adult. It shows that the general emotion model works better in the kindergarten group and performs similarly in the rest of the models.

**Book Length**

The elementary group performs the best in this category, with 0.475 in the Hits@5 metric. It also ranks first in the MRR metric result table. Middle-school and adult groups perform similarly in these two metrics. It confirms that younger children are more sensitive to book length.

**Cover image**

The result shows that younger children prefer brighter and more colorful images. Both metrics showed GDS_kindergarten > GDS_elementary > GDS_middle-school > GDS_adult. It suggests children are less attracted by the cover images when they grow up.

**Overall**

The result shows the performance differs in each age group. In most models, kindergarten performs outstandingly. This gives us an opposite conclusion from some research done in the past. In most models, the middle-school group has similar to the adult group. This suggests after growing up, children reading preferences changed towards adults.

## 4.2 Discussion

This section will discuss the result to find the answer to the research questions.

**RS performance on different age groups (RQ1)**

It is clear that RS has different performances across different ages. The popularity model performs well in the adult group. We think this is the correct signal since more adults provide feedback and reviews of the book and some purchase books based on the reviews. However, the kindergarten group result shows kindergarten has better accuracy at the top-5 recommendation list than the adult in rating models. This could cause by a few extremely popular books in the similar list. From the box plot, more outliers in the children group could lead to the higher Hits@5 result. Usually, children in kindergarten can not provide reviews or ratings by themselves. The researcher believes these ratings may come from their parents or supervisors. It is hard for children in kindergarten to purchase books by themselves; it is done mainly by their parents. This could be one of the reasons for explaining the distinctive data in the field. The description model has the worst performance in the middle-school group. The reason could be that middle school students prefer books with novel content. The topic in the kindergarten and elementary stages may be similar. Children prefer books with similar content as they had read before. It could also be the book in the same series. It is not expected that the children group still has the best performance in the emotion model. From the past study, the emotion model should work better for adults. But this experiment proposes a conflict signal. From the Milton et al.[17] research, books for children in kindergarten have higher distribution in the joy part and less distribution in other categories. The reason for this unexpected result is not apparent. It could be a bug in the experiment protocol. In the rest group, it is clear that the emotion model performs better when the children grow up. From the book length section, the younger group has better performance. It suggests that children in this age group prefer the same book length as they read before. Paper in the past[33] suggests the length has a practical impact on children's choice of book, and this model verifies this finding. It becomes less effective when the children grow up. Bright and colorful image has great attraction for kids in kindergarten. It verifies Milton et al.[17] research. After growing up, children prefer darker theme books. In conclusion, it is clear that RS performs differently between different age groups. When children grow up, the children-friendly model has less accuracy.

**Compare Traditional Model with Children Friendly Model(RQ2)**

In this experiment, the traditional model has a better interpretation than the children friendly model. However, the dataset's creation could be biased on the model based on popularity. The user creates the similar book list. The most voted book will be included in the list. That created a biased list based on popularity. It may be one reason the popularity model performs well in this experiment. The discussion about RQ1 mentioned the performance of the description similarity model. It could be the book in the same series. The children friendly model only uses one trait to create the recommendation models. It is possible to create a model using multiple traits, and it may achieve higher accuracy.

**Traits impact on RS(RQ3)**

The book length model has the most impact among the four models. It suggests children are more sensitive to the length of the book. It shows that children prefer books with a similar length for their next book. There is no sufficient data to support the emotion model. But in this research, the Text2Emotion package only analyzed the result from five perspectives. It may not be enough to analyze all the emotions in the description. The emotion analysis relies on the book's description, which may not represent the whole book's emotions. The description may write by other people instead the author themselves. The inaccuracy of the description could lead to a wrong emotion table in the analyzing procedure. Using other techniques to measure the similarity of two descriptions such as word2vec [38] is possible. This research also confirms the impact of the cover image on children. The result shows younger children prefer a book with brighter and more colorful images. It is evident in the kindergarten group. It suggests researchers can design an RS, mainly focusing on kindergarten groups with cover images. It may not work with other age groups. Research from Milton et al[17] shows contrast and entropy models work better for senior children. It is possible to use these two properties to create a new recommendation model with higher accuracy for elder children.

### 4.3 Limitation and Future Work

There are lots of limitations in this research. In general, the kindergarten performance is unusual. It has better performance in most models which is almost not possible. It could be a bug in the experiment code. Due to this project's time constraint, it is impossible to find the reason for this problem. The time problem also leads to an incomplete generation of the adult dataset. This research did not use the whole dataset from GoodReads, and it randomly selected 10% instances from the whole GoodReads dataset to generate GDS_Adult. This research did not try enough children friendly models, including literary elements, the cover image's contrast, and the entropy of the cover image. For the emotion model, only the description of the book is analyzed. It is unilateral text and may not write by the author. Analyzing the whole book text will lead to a more reasonable result to discuss. This research did not include the high school dataset due to insufficient data sources.

For future work, this research creates a fundamental step for evaluating RS for children with real-world reader-like data. The researcher is planning to examine more models introduced in Section 4.3. It is necessary to test these models with another dataset that contains chapters in the book. It helps the emotion model to analyze the book better and creates more reasonable results for analysis.

## 5 Responsible Research

No ethical-related aspects are involved in this research since the data are open-sourced and widely used in the community. The researcher notices that the children do not do most user-book interactions in the data source. Parents and guardians finished most of these interactions. This research does not use users' comments as a collaborative filtering method to prevent bias in children's book recommendations.

This research code will be open source on the GitHub[3]. However, the dataset will not be published. Due to the copyright declaration from the GoodReads[4] dataset creators, it is not allowed to redistribute the modified dataset. The repository in GitHub will contain the code to generate the GDS dataset. The code will be open source based on Apache License 2.0[5]. Distribution and modification are allowed, but warranty and liability will not be provided from this repository. Reproduce the result is possible by following Section 3 and the GitHub repository.

## 6 Conclusions

Book RS can help children to choose books and encourage them to read more books. In this research, it is not enough evidence proving that the children-friendly recommendation model works better than the traditional recommendation model. However, it is obvious RS performs differently when it targets different age groups. The cover image model performs better for younger children and the traditional recommendation model performs better for adult groups. The result indicates the potential for building an RS with the emotion model and the book length model together. However, it is not possible with the GoodReads dataset. From two models that analyze the cover image, it suggests children in kindergarten are more sensitive to brighter and more colorful cover images. It is possible to build an RS for kindergarten children based on the cover images.

[3]www.github.com/skyloveqiu/children_rs

[4]https://sites.google.com/eng.ucsd.edu/ucsdbookgraph/home

[5]https://www.apache.org/licenses/LICENSE-2.0

## References


[1] **Khalifeh, Sewar and Amjed A. Al-Mousa**. "A Book Recommender System Using Collaborative Filtering Method". In: *International Conference on Data Science, E-learning and Information Systems 2021* (Apr. 2021). DOI: 10.1145/3460620.3460744. URL: http://dx.doi.org/10.1145/3460620.3460744.

[2] **Billsus**. "Content-Based Recommendation Systems". In: *The Adaptive Web* (2007), pp. 325–341. DOI: 10.1007/978-3-540-72079-9{_}10/. URL: http://dx.doi.org/10.1007/978-3-540-72079-9_10.

[3] **Frankowski, Herlocker, and Sen**. "Collaborative Filtering Recommender Systems". In: *The Adaptive Web* (Jan. 2007), pp. 291–324. DOI: 10.1007/978-3-540-72079-9\{_}9. URL: http://dx.doi.org/10.1007/978-3-540-72079-9_9.

[4] **Smith, Brent and Greg Linden**. "Two Decades of Recommender Systems at Amazon.com". In: *IEEE Internet Computing* 21.3 (May 2017), pp. 12–18. DOI: 10.1109/mic.2017.72. URL: http://dx.doi.org/10.1109/mic.2017.72.

[5] **Jannach, Dietmar and Michael Jugovac**. "Measuring the Business Value of Recommender Systems". In: *ACM Transactions on Management Information Systems* 10.4 (Dec. 31, 2019), pp. 1–23. DOI: 10.1145/3370082. URL: http://dx.doi.org/10.1145/3370082.

[6] **Gomez-Uribe, Carlos A. and Neil Hunt**. "The Netflix Recommender System". In: *ACM Transactions on Management Information Systems* 6.4 (Jan. 14, 2016), pp. 1–19. DOI: 10.1145/2843948. URL: http://dx.doi.org/10.1145/2843948.

[7] **Wang, Zhi Hui and De Zhi Hou**. "Research on Book Recommendation Algorithm Based on Collaborative Filtering and Interest Degree". In: *Wireless Communications and Mobile Computing* 2021 (July 10, 2021). Ed. by Balakrishnan Nagaraj, pp. 1–7. DOI: 10.1155/2021/7036357. URL: http://dx.doi.org/10.1155/2021/7036357.

[8] **Hofferth, Sandra L. and Ui Jeong Moon**. "Cell phone use and child and adolescent reading proficiency." In: *Psychology of Popular Media Culture* 1.2 (Apr. 2012), pp. 108–122. DOI: 10.1037/a0027880. URL: http://dx.doi.org/10.1037/a0027880.

[9] **Hofferth, Sandra L. and Ui Jeong Moon**. "Electronic Play, Study, Communication, and Adolescent Achievement, 2003-2008". In: *Journal of Research on Adolescence* 22.2 (Nov. 21, 2011), pp. 215–224. DOI: 10.1111/j.1532-7795.2011.00770.x. URL: http://dx.doi.org/10.1111/j.1532-7795.2011.00770.x.

[10] **Hosokawa, Rikuya and Toshiki Katsura**. "Association between mobile technology use and child adjustment in early elementary school age". In: *PLOS ONE* 13.7 (July 25, 2018). Ed. by Danuta Wasserman, e0199959. DOI: 10.1371/journal.pone.0199959. URL: http://dx.doi.org/10.1371/journal.pone.0199959.

[11] **Ludewig, Ulrich et al.** "COVID-19 Pandemic and Student Reading Achievement: Findings From a School Panel Study". In: *Frontiers in Psychology* 13 (May 2022). DOI: 10.3389/fpsyg.2022.876485. URL: http://dx.doi.org/10.3389/fpsyg.2022.876485.

[12] **Patrinos, Harry Anthony, Emiliana Vegas, and Rohan Carter-Rau**. "An Analysis of COVID-19 Student Learning Loss". In: *Policy Research Working Papers* (May 2022). DOI: 10.1596/1813-9450-10033. URL: http://dx.doi.org/10.1596/1813-9450-10033.

[13] **Mondal, Anwesha and Manish Kumar**. "A study on Internet addiction and its relation to psychopathology and self-esteem among college students". In: *Industrial Psychiatry Journal* 27.1 (2018), p. 61. DOI: 10.4103/ipj.ipj_61_17. URL: http://dx.doi.org/10.4103/ipj.ipj_61_17.

[14] **Mak, Hei Wan and Daisy Fancourt**. "Reading for pleasure in childhood and adolescent healthy behaviours: Longitudinal associations using the Millennium Cohort Study". In: *Preventive Medicine* 130 (Jan. 2020), p. 105889. DOI: 10.1016/j.ypmed.2019.105889. URL: http://dx.doi.org/10.1016/j.ypmed.2019.105889.

[15] **Ng, Yiu Kai**. "CBRec: a book recommendation system for children using the matrix factorisation and content-based filtering approaches". In: *International Journal of Business Intelligence and Data Mining* 16.2 (2020), p. 129. DOI: 10.1504/ijbidm.2020.104738. URL: http://dx.doi.org/10.1504/ijbidm.2020.104738.

[16] **Gómez, Charisi, and Chaudron**. "Evaluating recommender systems with and for children: towards a multi-perspective framework". In: *Perspectives@ RecSys* (2021).

[17] **Milton, Ashlee et al.** ""Don't Judge a Book by its Cover": Exploring Book Traits Children Favor". In: *RecSys 2020: Fourteenth ACM Conference on Recommender Systems, Virtual Event, Brazil, September 22-26, 2020*. Ed. by Rodrygo L. T. Santos et al. ACM, 2020, pp. 669–674. DOI: 10.1145/3383313.3418490. URL: https://doi.org/10.1145/3383313.3418490.

[18] **Ricci, Francesco, Lior Rokach, and Bracha Shapira**. "Recommender Systems: Techniques, Applications, and Challenges". In: *Recommender Systems Handbook* (Nov. 22, 2021), pp. 1–35. DOI: 10.1007/978-1-0716-2197-4_1. URL: http://dx.doi.org/10.1007/978-1-0716-2197-4_1.

[19] **Gomez-Uribe, Carlos A. and Neil Hunt**. "The Netflix Recommender System". In: *ACM Transactions on Management Information Systems* 6.4 (Dec. 28, 2015), pp. 1–19. DOI: 10.1145/2843948. URL: http://dx.doi.org/10.1145/2843948.

[20] **Sarwar, Badrul et al.** "Item-based collaborative filtering recommendation algorithms". In: *Proceedings of the tenth international conference on World Wide Web - WWW '01* (2001). DOI: 10.1145/371920.372071. URL: http://dx.doi.org/10.1145/371920.372071.

[21] **Burke, Robin**. "Hybrid Recommender Systems: Survey and Experiments". In: *User Modeling and User-Adapted Interaction* 12.4 (2002), pp. 331–370. DOI: 10.1023/a:1021240730564. URL: http://dx.doi.org/10.1023/a:1021240730564.

[22] **Isinkaye, F.O., Y.O. Folajimi, and B.A. Ojokoh**. "Recommendation systems: Principles, methods and evaluation". In: *Egyptian Informatics Journal* 16.3 (2015), pp. 261–273. ISSN: 1110-8665. DOI: https://doi.org/10.1016/j.eij.2015.06.005. URL: https://www.sciencedirect.com/science/article/pii/S1110866515000341.

[23] **Hossin, Mohammad and Sulaiman M.N**. "A Review on Evaluation Metrics for Data Classification Evaluations". In: *International Journal of Data Mining Knowledge Management Process* 5 (Mar. 2015), pp. 01–11. DOI: 10.5121/ijdkp.2015.5201.

[24] **Ge, Mouzhi, Carla Delgado-Battenfeld, and Dietmar Jannach**. "Beyond accuracy". In: *Proceedings of the fourth ACM conference on Recommender systems* (Sept. 26, 2010). DOI: 10.1145/1864708.1864761. URL: http://dx.doi.org/10.1145/1864708.1864761.

[25] **Shani, Guy and Asela Gunawardana**. "Evaluating Recommendation Systems". In: *Recommender Systems Handbook* (Oct. 5, 2010), pp. 257–297. DOI: 10.1007/978-0-387-85820-3_8. URL: http://dx.doi.org/10.1007/978-0-387-85820-3_8.

[26] **Fayyaz, Zeshan et al.** "Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities". In: *Applied Sciences* 10.21 (Nov. 2, 2020), p. 7748. DOI: 10.3390/app10217748. URL: http://dx.doi.org/10.3390/app10217748.

[27] **Craswell, Nick**. "Mean Reciprocal Rank". In: *Encyclopedia of Database Systems*. Ed. by Ling Liu and M. Tamer Özsu. New York, NY: Springer New York, 2018, pp. 2217–2217. ISBN: 978-1-4614-8265-9. DOI: 10.1007/978-1-4614-8265-9_488. URL: https://doi.org/10.1007/978-1-4614-8265-9_488.

[28] **Pera, Maria Soledad and Yiu-Kai Ng**. "What to read next?" In: *Proceedings of the 7th ACM conference on Recommender systems* (Oct. 12, 2013). DOI: 10.1145/2507157.2507181. URL: http://dx.doi.org/10.1145/2507157.2507181.

[29] **Milton, Ashlee et al.** "StoryTime: Eliciting Preferences from Children for Book Recommendations". In: *Proceedings of the 13th ACM Conference on Recommender Systems*. RecSys '19. Copenhagen, Denmark: Association for Computing Machinery, 2019, pp. 544–545. ISBN: 9781450362436. DOI: 10.1145/3298689.3347048. URL: https://doi.org/10.1145/3298689.3347048.

[30] **Pera, Maria Soledad and Yiu-Kai Ng**. "Automating Readers' Advisory to Make Book Recommendations for K-12 Readers". In: *Proceedings of the 8th ACM Conference on Recommender Systems*. RecSys '14. Foster City, Silicon Valley, California, USA: Association for Computing Machinery, 2014, pp. 9–16. ISBN: 9781450326681. DOI: 10.1145/2645710.2645721. URL: https://doi.org/10.1145/2645710.2645721.

[31] **Hakim, Ari Aulia et al.** "Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach". In: *2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE)* (Oct. 2014). DOI: 10.1109/iciteed.2014.7007894. URL: http://dx.doi.org/10.1109/iciteed.2014.7007894.

[32] **Tkalcic, Marko, Andrej Košir, and Jurij F. Tasic**. "Affective recommender systems: The role of emotions in recommender systems". In: *CEUR-WS.org eBooks* (Jan. 1, 2011).

[33] **Jalilehvand, Maryam**. "The Effects of Text Length and Picture on Reading Comprehension of Iranian EFL Students". In: *Asian Social Science* 8.3 (Feb. 28, 2012). DOI: 10.5539/ass.v8n3p329. URL: http://dx.doi.org/10.5539/ass.v8n3p329.

[34] *2018 reform of EU data protection rules*. European Commission. May 25, 2018. URL: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf (visited on 06/17/2019).

[35] **Wan, Mengting et al.** "Fine-Grained Spoiler Detection from Large-Scale Review Corpora". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Florence, Italy: Association for Computational Linguistics, July 2019, pp. 2605–2610. DOI: 10.18653/v1/P19-1248. URL: https://aclanthology.org/P19-1248.

[36] **Wan, Mengting and Julian J. McAuley**. "Item recommendation on monotonic behavior chains". In: *Proceedings of the 12th ACM Conference on Recommender Systems, RecSys 2018, Vancouver, BC, Canada, October 2-7, 2018*. Ed. by Sole Pera et al. ACM, 2018, pp. 86–94. DOI: 10.1145/3240323.3240369. URL: https://doi.org/10.1145/3240323.3240369.

[37] **Hall, Barbara I., Sherywn P. Morreale, and James L. Gaudino**. "A survey of the status of oral communication in the K-12 public educational system in the United States". In: *Communication Education* 48.2 (Apr. 1999), pp. 139–148. DOI: 10.1080/03634529909379161. URL: http://dx.doi.org/10.1080/03634529909379161.

[38] **Mikolov, Tomas et al.** "Efficient Estimation of Word Representations in Vector Space". In: *arXiv: Computation and Language* (Jan. 16, 2013). URL: http://export.arxiv.org/pdf/1301.3781.