**Content and collaborate based Sinhala Book Recommendation System**

P. N. C. Perera

University of Colombo School of Computing, Colombo 07, Sri Lanka

[nc21perera@gmail.com](mailto:nc21perera@gmail.com)

Prof. M.G.N.A.S. Fernando

University of Colombo School of Computing, Colombo 07, Sri Lanka

[nas@ucsc.cmb.ac.lk](mailto:nas@ucsc.cmb.ac.lk)

**Abstract**

The saying "Reading makes a man perfect" tells us how important it is to gain knowledge by reading different things like books, articles on the internet, newspapers, magazines, or even simple pieces of paper. This article talks about how reading a lot can have a big impact on people. It says that people who read regularly usually have a better understanding of life, make smarter choices, and handle difficult situations better. The traditional way of picking books, like choosing randomly or going by recommendations, can be hard, especially as libraries have more and more books. The COVID-19 pandemic and distance can make it even harder for people to find the books they want. This situation shows that there is a need for a better way for people to borrow books from the library. While many online platforms use systems that suggest products, these usually focus on making more sales and not necessarily on what the user really likes. This article talks about the lack of a system that suggests Sinhala books and suggests a new idea. The proposed solution combines both content and collaboration methods to create a hybrid model for recommending Sinhala books. This system aims to help online users find interesting and relevant books without wasting too much time and money. It not only benefits readers but also gives useful information to authors, helping them understand what readers like and adjust their writing to match popular interests. This recommendation system has the potential to change the way Sinhala books are discovered and chosen, making it a useful tool for both readers and authors in the digital age.

***Keywords:*** Sinhala Books, Collaborative filtering, Content based filtering, Hybrid model, Recommendation system

**1. Introduction**

With the increase of online shopping items, the need for giving the confidence in buying products has been increased. Many ecommerce companies like Amazon, Netflix, Spotify utilizing the power of recommendation system by implementing in their website to boost their sales. Even machine learning technique can be used to solve so many problems in real world, making product recommendation is widely known application of machine learning.

There are many recommendation systems have been implemented for many domains such as movies, books, e commerce items and songs. In this research the author is implementing a book recommendation system which help readers to find similar books or books read by other users who has similar preference as you. Since most applications have been implemented for English books and there is no system for Sinhala books, this research is focusing on implementing a book recommendation for Sinhala books.

People used to read from their childhood. As per the research, it says, People who read a lot tend to know more about life and are smarter when making decisions and handling difficult situations. (Marappan, 2022). In today’s world, time has more value and the researchers have no much time to spend on searching for the right articles according to their research domain. (Murali et al., 2019). Book readers usually select books by reading some random pages or asking someone to recommended any book. When reading that book, if he finds that the book is not interesting, he will not read any book after that. therefore, it is better to suggest books that he is interested in. With the increase in library collections, it is diﬃcult for readers to quickly ﬁnd the books they want. It is also diﬃcult for readers to ﬁnd Sinhala books of interest in a short period of time in the face of various bibliographies. Therefore, the user experience of the traditional library borrowing method is poor.(Dhanda and Verma, 2016) Due to the Covid-19 pandemic situation and the geographical barriers also it becomes a tremendous challenge for readers (Sarma et al., 2021) to find a relevant book as they do not like to go out and spend time searching books of their preference. Even the pandemic period is over it is better to be prepared to face such situation in future.

**2. Literature Review**

Several researches have been conducted related to Book recommendation systems and most of them have used machine learning techniques to build the model which is used to generate the list of items that user prefer.

In a study by (Sarma et al., 2021) researchers created a book recommendation system for online users. They used clustering to rate and suggest books based on similarity, utilizing data from Kaggle's Good readers book repository. The system identified and excluded potentially boring books through a classifier. Evaluation metrics included precision, sensitivity, specificity, and F1 score, along with a graphical accuracy representation through a receiver operating characteristic (ROC) curve. Future work aims to propose a system for recommending online courses using Convolutional Neural Network (CNN) technology.

The research (Wadikar et al., 2020) introduces a platform using a Convolutional Neural Network (CNN) for book recommendations through text processing and image classification. Text input from users is processed, and a dataset is collected by web scraping Amazon and Flipkart. In image classification, users upload book cover images for results. Cosine similarity is used to find related books. Advantages include no need for feature engineering, efficient handling of unstructured data, and quick access to highly-rated books. Evaluation and validation processes are lacking in the research.

The study by (Shah, 2019) created an e-commerce application employing collaborative filtering algorithms. Users could input ratings or sentences, and the system used natural language processing to calculate ratings based on sentence nature. The research delved into challenges like Scalability, Sparsity, Security, Cold start, and veracity in recommendation systems. Various methods were discussed, including clustering, classification, and item-based collaboration. The study used the "goodbooks10k" dataset on Kaggle, experimented with Python, and assessed accuracy using Mean Absolute Error (MEA).

Top of Form

**3. Theoretical Concepts**

***3.1 Content based***

A content-based book recommendation system suggests books to users based on the characteristics of books they have already liked. It works by analyzing the content or features of books, such as genres, authors, and keywords, to find similarities with the user's preferences. For example, if a user enjoys mystery novels with a specific author, the system will recommend other mystery books by the same author or within the same genre. This approach relies on understanding the intrinsic qualities of books and matching them to the user's known preferences, making it a personalized recommendation system that doesn't require information about other users.

***3.2 Collaborate based***

A collaborative-based book recommendation system works by analyzing the preferences and behaviors of a group of people to suggest books. Instead of relying on individual preferences alone, it considers the collective tastes and choices of a community. If people with similar reading habits enjoyed a particular book, the system might recommend that book to others with comparable tastes. This approach leverages the wisdom of the crowd, making recommendations based on the shared preferences of a community rather than focusing solely on an individual's likes and dislikes. It's like getting book suggestions from a group of friends who have similar reading interests, making the recommendations more likely to match your taste.

***3.3 Sentiment Analysis***

Sentiment analysis is like teaching computers to understand feelings in written text. Imagine you write something online, like a review or a tweet, and the computer reads it to figure out if you're happy, sad, angry, or neutral. It's like a digital detective for emotions! This helps businesses know how people feel about their products or services, and it also allows social media platforms to filter out negative or harmful content. So, sentiment analysis is a smart way for computers to catch the vibe of what people are saying online.

***3.4 Artificial Neural Network***

Artificial Neural Network (ANN) is kind of a virtual brain inspired by the way our own brains work. It's a computer system designed to learn and make decisions, just like humans do. In this network, there are nodes, or artificial neurons, that are connected in layers. These connections have weights that determine the strength of the relationship between the neurons. During training, the network learns from examples, adjusting these weights to improve its ability to make predictions or classifications. It's like teaching the network to recognize patterns and make sense of information. Once trained, the ANN can take in new data and use its learned knowledge to make predictions or decisions. So, in a nutshell, an Artificial Neural Network is a computer system that learns and thinks in a way inspired by the human brain to solve problems or make decisions.

**4. Methodology**

***4.1 Data Collection***

Since there are no data set available in the data set provides like Kaggle, a google form was shared among book readers groups in order to collect considerable amount of dataset for the research. 845 users contributed by filling the form with 5 books each and total of 4225 records could be collected. There are more than 300 popular books were considered for the research and each book details were collected via online book stores and saved in the database in order to display the book details once a book is selected by a user in the application.

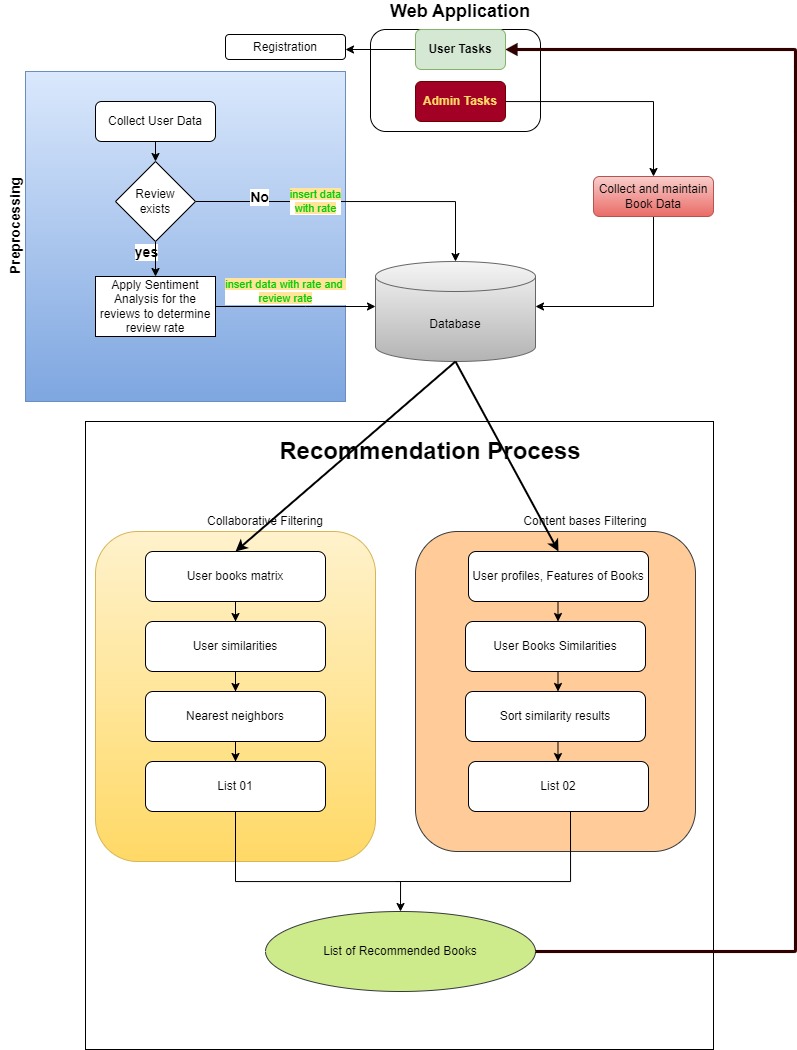
***4.2 Preprocessing***

In the google sheet shared to the user, there were two significant fields name rate and review. The rate could be selected in a range from one to ten and based on the provided rate a user can enter a review considering why you have given such rate for the selected book. Since a user can select a good rate and provide a negative review, a mechanism had to be used to overcome the problem. That is applying sentiment analysis to the review and generate a new rate based on the review provided. The final rate was considered by calculating the mean value or user given rate and review rate. Additionally, two separate models for content based and collaborate based have been created to predict the final book list.

***4.4 Technology stack***

The recommendation system was implemented using Python and popular libraries such as python Flask, Google translator in python, nltk and scikit-learn.Flask Python library with html, css and bootstrap ware used to implement the web application and scikit-learn, nltk, google translator in python were used in preprocessing and main business logics. Mysql workbench was used as main database to store and maintain the data

***4.3 Architectural Diagram***

.

In this system, data is first collected and then organized into three tables: Book\_Details, User\_Details, and Rate\_Details. User information is stored in User\_Details, while Book\_Details contains additional book details. Ratings given by users for books are stored in Rate\_Details, and if there's a review, a separate review rate is assigned after sentimental analysis and included in Rate\_Details. The final calculation involves taking the mean value of review and normal rates. Data preprocessing is essential before recommendation, removing null rate values and eliminating instances of maximum rates for a single book.

The recommendation process involves applying algorithms, with Collaborative Filtering being a primary method. This method recommends books based on similarities found from other users. User-based collaborative algorithms involve building a matrix of users and books with ratings, computing cosine similarity, and generating a list of recommended books. Content-based approaches focus on book descriptions and user preferences, creating a user preference vector based on past ratings. Hybrid methods, combining collaborative and content-based filtering, are commonly used for more effective book recommendations. The final recommended book list is generated by displaying the common books from collaborative and content-based filtering.

fig 1: Architect Diagram

The system implementation also addresses challenges, such as the increasing number of research papers. Efficient searching and filtering mechanisms, as mentioned in (Murali et al., 2019), are crucial to save researchers' time. The paper highlights the use of algorithms to recommend quality research papers based on users' queries and preferences.

In conclusion, this system involves collecting, categorizing, and processing data to recommend books efficiently. It employs Collaborative Filtering, Content-Based Filtering, and a Hybrid approach for effective recommendations. The process includes data preprocessing, algorithm application, and a final recommendation list generation, enhancing the overall user experience in finding relevant books.

**5. Results and Discussion**

***5.1 Evaluate Collaborate Filter - Mean Absolute Error (MAE)***

Even though there are many methods available to evaluate the system, the paper (Raval and Khedkar, 2019) used item based collaboration filtering recommendation system and the Root Mean Square Error (RMSE) and Mean Absolute Error (MSE) methods were used for evaluation respectively. In the final results, their proposed method outperforms all the state-of-art methods. To align with the above research, the system implemented by (Shah, 2019) also used Mean Absolute Error (MSE) for the evaluation.

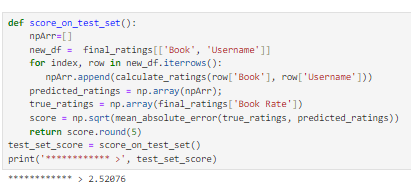


fig 2: MAE value

Fig. 2 shows the calculated Mean Absolute Error value as 2.52076 which is closer to zero. we received a good score for the accuracy for the collaboration model.

***5.2 Evaluate Content based Filter – Artificial Neural Network (ANN)***

The above MAE was used to calculate the accuracy of the approach which has a numeric field in our case ‘Book Rate’. The data set used for collaborative filtering have the rate field. But as per the data set of books having tags does not have any numeric field and therefore MAE cannot be applied for Content based filtering. But since the accuracy gives the correctness of the implemented application, a different approach which only works with text should be used. We integrate Artificial Nural Network (ANN) for the application and predict the recommended book list. The results can be compared and accuracy can be calculated with implemented Content based model.

*5.2.1* Artificial Nural Network (ANN)

The Artificial Nural Network is connected network which takes an input value and computes the desired output. The book with tags dataset can be considered as input data and recommended book list is the output. The reason behind selecting the ANN is it’s not just giving the recommended books but also compare the results with accuracy percentage.

Following table shows the accuracy for the selected book. There we generate the results using Artificial Neural Network and checked what percentage of books were generated by using content-based filtering. Most books listed in content based and calculate all together we received 80% of the accuracy for content-based filtering.

|  |  |  |
| --- | --- | --- |
| **User Selected Book** | **Number of books recommended by Content based and are listed in the list recommended by ANN out of 20 Books** | **Accuracy** |
| Oliver Twist - M. M. Piyawardana | 17 | 85% |
| Hari Puduma Iskole - Leelananda Gamachchi | 19 | 95% |
| Bhayanaka Miniha - Chandana Mendis | 16 | 80% |
| Rathu Rosa - Kumara Karunarathna | 14 | 70% |
| Sanda Wiyaruwa - Bhadraji Mahinda Jayathilaka | 13 | 65% |
| Iti Pahan - Sumithra Rahubadda | 13 | 65% |
| Gahanu Lamayi - Karunasena Jayalath | 20 | 100% |
| 105 - Dileepa Jayakody | 16 | 80% |
| Apuru Iskole Apuru Dawas - Sudath Rohan | 16 | 80% |
| Bindunu Bilinda - Dileepa Jayakody | 13 | 65% |
| Emily 01 - Manel Jayanthi Gunasekara | 18 | 90% |
| Anne 01 (Arabe Gedara Anne) - Premasiri Mahingoda | 17 | 85% |
| **Total Average** |  | **80%** |

Table 1: Accuracy of Content based filter

***5.3 Online Survey for final result***

This study will use 32 test subjects A questionnaire was prepared to determine user satisfaction and the quality of the suggested book list for the users. Five-point Likert-scale survey questions were asked. Likert scales have become an essential survey tool to get feedback on a person’s opinion or attitude regarding an item. It ranges from polar opposites to complete satisfaction to complete dissatisfaction. Questions were structured to be asked under the categories of accuracy, familiarity, novelty of the book recommendations, and interactivity of the system. An optional question was asked if the user wanted to give any suggestions or feedback for further improving the system. This questionnaire determines whether the implemented system has met the objectives and met the user's requirements and needs.

|  |  |  |  |
| --- | --- | --- | --- |
| **Question** | **Positive Count** | **Total** | **Percentage** |
| I think a system is required to find similar books based on users’ preference or content as I am struggled finding new books similar to books I have read and interested. | 29 | 32 | 90,63 |
| The implemented system accurately recommends books | 31 | 32 | 96,88 |
| The system helped me to find new books | 31 | 32 | 96,88 |
| It is easy to navigate and use the system | 29 | 32 | 90,63 |
| Overall, I’m satisfied with the recommender system | 31 | 32 | 96,88 |
| I would recommend this Sinhala Book Recommendation system to others. | 31 | 32 | 96,88 |

Table 2: Feedback of online survey

**6. Conclusion**

This solid waste generation, providing decision makers, useful information for waste management policy development.

**7. References**