

## Model name

BookHarbor Recommendation System

## Model date and version

Built: August 2024

Updates: No updates have been made since the initial version. Users will be informed via email if updates are released.

## Overview Model

BookHarbor is a recommendation system designed to provide personalized book suggestions for college students in Kathmandu. By leveraging advanced machine learning algorithms, the system offers tailored book recommendations based on user preferences, reading patterns, and interests. The system uses a combination of collaborative filtering, popularity-based filtering, and hybrid approaches to analyze large datasets, including metadata, user reviews, and reading trends.

## Questions or comments

Please send any questions to:

[team@datascience\\_ioa.org](mailto:team@datascience_ioa.org)

## Primary intended users

College students in Kathmandu

Academic institutions

## Primary intended uses

To provide personalized book recommendations to college students

To enhance the reading experience by suggesting books that match users' interests and reading habits

## Out of scope uses

Not intended for real-time decision-making impacting financial or critical outcomes

Not suitable for non-academic or commercial book recommendation purposes

## Limitations

The system is optimized for the student population in Kathmandu and may not generalize well to other demographics.

Recommendations may not reflect real-time changes in user preferences as the model is not updated in real-time.

Limited access to diverse and international book collections.

## Metrics

**Precision:** Measures the accuracy of positive recommendations.

**Recall:** Measures the ability to identify all relevant recommendations.

**Accuracy:** Overall correctness of the model's predictions.

## Training and evaluation data

**Dataset:** Includes book metadata, user reviews, and reading patterns from college students in Kathmandu.

**Data Size:** Approximately 10,000 records.

**Imbalance:** The dataset may have an imbalance due to varying levels of user engagement and reading habits.

## Quantitative analysis

### Accuracy and reliability

The BookHarbor recommendation system has an accuracy level of 85%, meaning it was able to place books into the correct category 85 times out of a hundred. The data set was imbalanced, with only 15% of books receiving high ratings in the training data set, resulting in potential biases in accuracy rates. The training set of 10,000 books should be adequate in size and given the number of parameters to provide reliable results.

### Precision

Precision measures the accuracy of the positive (highly recommended) label. A low score on precision indicates a high number of false positives. The precision rate for the BookHarbor system was 80%, meaning that 20% of the books labeled as highly recommended were not rated as such by users.

## Recall

Recall measures the ability of the algorithm to identify all potential positives (highly recommended books that users liked). A low score on recall indicates a high number of false negatives. The recall rate for the BookHarbor system was 93%, meaning that only 7% of the highly recommended books were missed. These results are within our benchmark criteria, but the limitations should be considered when making decisions.

## Ethical considerations

The intended use of the BookHarbor recommendation system is for enhancing the reading experience of college students in Kathmandu by providing personalized book suggestions. No personally identifiable information (PII) was included in the training and test data sets. The role that any sensitive information (such as ethnicity, gender, or religious affiliation) may have played in the recommendation process was not investigated.

The system was developed and evaluated using data up until August 2024. Further work is needed to identify whether the recommendation algorithm continues to generalize effectively to new users over time. Generalization refers to the ability of an algorithm that was trained on one specific data set to work equally efficiently when applied to a new data set.

The model does not capture information on potential users with limited engagement with digital reading platforms, and the findings should not be generalized to that group.

## Feedback

For questions or comments, please contact:

- **Email:** [support@bookharbor.edu](mailto:support@bookharbor.edu)
- Users are encouraged to share feedback to help improve the system's performance and accuracy.

## Additional notes and any other relevant factors

The model's performance should be periodically evaluated to ensure it continues to meet the needs of the intended users.

Consideration of integrating more diverse datasets to enhance the recommendation quality for a broader audience.