

Apache Mahout based Book Recommendation System



Abhishek Verma, V Nallarasan

Abstract: E-Commerce websites plays an important role in an individual's life as it serves as the medium for online shopping with a huge audience. With the commencement of the pandemic due to novel coronavirus, the involvement of E-Commerce websites for shopping has drastically increased or more precisely it remains as the only medium to shop. With the increasing demand for online shopping on E-Commerce websites, the role of the Recommendation System has also become vital as it accomplishes the goal to make Personalized Recommendation for users. In this paper, we set out Apache Mahout-based Book Recommendation System to help recommend books to users. With this paper, we have described our project that recommends books to users on the basis of the user's prior experience of purchase. The platform utilizing this recommendation system is developed using Spring Framework as a part of our project. The dataset used in the process is taken from Kaggle. Dataset has ratings for various books given by users. As a part of the User-based Collaborative Filtering recommendation technique, Euclidean Distance Similarity is used as a similarity measure along with Nearest N User Neighborhood and Generic User-Based Recommender to give quality results as compared to the existing system. To get the best quality recommendation we have obtained an evaluation score of 0.5 for Euclidean Distance Similarity.

Keywords: Apache Mahout, Book Recommendation, Collaborative Filtering, Machine Learning, Spring Framework, Web Application.

I. INTRODUCTION

With the commencement of the pandemic due to novel coronavirus, the involvement of E-Commerce websites for shopping has drastically increased or more precisely it remains as the only medium to shop. With such an increase in the demand for online shopping platforms, it is important that users get the best quality recommendations based on the purchases made in past. Our project uses Apache Mahout, a Machine Learning framework to implement Collaborative Filtering based recommendation system to recommend books to users on an online book shopping platform developed using Spring Framework. The focus of the project is to help customers get the best quality recommendations and avail the best offerings from an E-Commerce platform when the online mode is the most suited way for shopping. The recommendation system is of immense value in a current

situation not only restricted to books but also for a vast variety of products including various essentials required during the pandemic.

II. LITERATURE REVIEW

The previous researcher's Machine Learning techniques and approaches for recommending items have been quite successful. Saikat Bagchi at IIT Kharagpur [1] has analyzed and compared various similarity measures which is an important aspect to make recommendations using Collaborative Filtering. As a result of the study conducted it is concluded that Euclidean Distance Similarity performs very well as compared to other similarity measures which are being used in our project. [2] Dilek Tapucu, Seda Kasap, Fatih Tekbacak have shown combined solution results using various similarity measures. They have described that Pearson Correlation which is user-based CF algorithm has a better performance. They have also proved that combined user and item-based CF algorithms can perform better in some scenarios. [3] Johnpaul C I, Neetha Susan Thampi, Dr. Senthil Kumar Thangavel have concluded that Apache Mahout can handle a large amount of structured data which is being used in our project. [4] Ananya Agarwal, S. Srinivasan have used Pearson Correlation Similarity as a similarity measure in the Collaborative Filtering technique for building a Movie Recommendation System. [5] Abhilasha Sase, Kritika Varun, Sanyukta Rathod, Deepali Patil have proposed that a hybrid recommendation system is more accurate and efficient as it combines the features of various recommendation techniques [6]. In summary, there are many existing works around Collaborative Filtering Based Recommendation Systems and most of them use data being provided by the users on E-Commerce sites to recommend items to them [7]. There are works existing around Book Recommendation Systems as well [8]. Due to the sudden increase in demand of online shopping, it is vital that customers get the best quality recommendations and avail best offerings from these sites for the books or any other items that they purchase [9].

III. DATASET DESCRIPTION

The dataset for Book Recommendation was taken from Kaggle, a repository for data and scientific papers related to diverse fields and can be accessed online [10]. This dataset consists of three columns: the first column as User Id, the second column as Book Id, and the third column as Book Rating. Every user will have a unique id and can give ratings ranging from '1' to '5' with the higher value being the best rating for every book having a unique book id [11]. The dataset has 981756 entries with users giving ratings to various books ranging from '1' to '5'.

Manuscript received on September 08, 2021.

Revised Manuscript received on February 11, 2022.

Manuscript published on February 28, 2022.

* Correspondence Author

Abhishek Verma*, Department of Information Technology, SRM Institute of Science and Technology, Chennai (Tamil Nadu), India. Email: abhishek.verma4607@gmail.com

V Nallarasan, Assistant Professor, Department of Information Technology, SRM Institute of Science and Technology, Chennai (Tamil Nadu), India. Email: nallarav@srmist.edu.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)



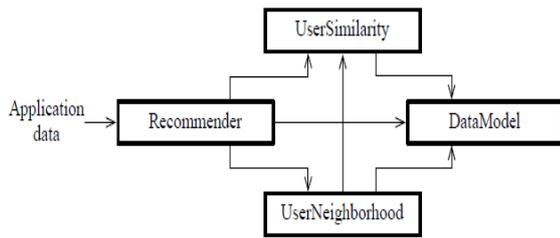


Fig. 13. Interaction between components in Apache Mahout User Based Recommender

VII. PERFORMANCE OF FINAL METHOD

As we are recommending books to users based on the User-based Collaborative Filtering Algorithm. The similarity measure plays a vital role in deciding the quality of recommendations being made to the user. Based on the experiment conducted we arrived at a conclusion that Euclidean Distance Similarity (EDS) has the lowest evaluation score of 0.5 as compared to other similarity measures and thus provides the best quality recommendation.

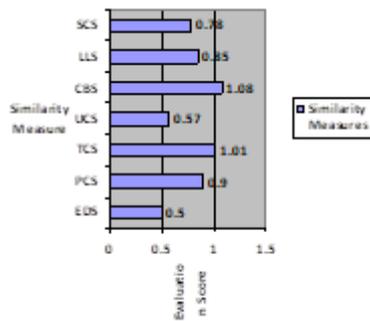


Fig. 14. Evaluation Score of various Similarity Measures calculated based on MAE

VIII. CONCLUSION AND FUTURE WORK

In our model, we are using Euclidean Distance Similarity as the similarity measure to produce best quality recommendations because it has a low evaluation score as shown in “Fig. 14”. Therefore, we conclude that the Euclidean Distance Similarity has been found to be the most appropriate similarity measure for providing quality recommendations. We bring forward this substructure to assist users to shop online by getting quality recommendations. This same methodology can be used for recommending a variety of products to users. For our future work, we would like to look at how the proposed methodology works with datasets containing a variety of products including scenarios where all the products are not rated.

REFERENCES

1. Saikat Bagchi (2015) - “Performance and Quality Assessment of Similarity Measures in Collaborative Filtering Using Mahout”: IIT Kharagpur, India.
2. Ananya Agarwal, S. Srinivasan (2020) - “Movie Recommendation System” IRJET: Department of Computer Science, Galgotias University, Uttar Pradesh.
3. Dilek Tapucu, Seda Kasap, Fatih Tekbacak (2012) - “Performance Comparison of Combined Collaborative Filtering Algorithm for Recommender Systems” IEEE: Department of Computer Engineering, Izmir Institute of Technology, Urla, Izmir, Turkey and Faculty of

Engineering and Natural Sciences, Sabanci University, Orhanli, Tuzla, Istanbul, Turkey.

4. Taner Arsan, Efekan Koksul, Zeki Bozkus (2016) - “Comparison of Collaborative Filtering Algorithm with various Similarity Measures for Movie Recommendation” IJCSEA: Department of Computer Engineering, Kadir Has University, Istanbul, Turkey.
5. Abhilasha Sase, Kritika Varun, Sanyukta Rathod, Deepali Patil (2015) - “A Proposed Book Recommender System” IJARCCCE: Sinhgad Institute of Science and Technology, Pune, India.
6. Tarun Bhatia, Upendra Chaurasia (2016) - “User-Based Collaborative Filtering Recommendation System using Apache Mahout”, Gregoria Institute of Science and Technology.
7. Dr. Senthil Kumar Thangavel, Neetha Susan Thampi, Johnpaul C I (2013) - “Performance Analysis of Various Recommendation Algorithms Using Apache Hadoop and Mahout”, IJSER.
8. S. Kanetkar, Akshay Nayak, S. Swamy, G. Bhatia (2014) - “Web-based personalized hybrid book recommendation system”, ICAETR.
9. <https://mahout.apache.org/>
10. S. G. Walunj, K Sadafale, “An online recommendation system for e-commerce based on Apache Mahout framework”, 2017 ACM SIGMIS International Conference on Computers and People Research, pp 153-158, 2013.
11. Z D Zhao, M. Shang,” User-Based collaborative filtering recommendation algorithms on Hadoop”, Proc. Of Third International Workshop on Knowledge Discovery and Data Mining, pp. 478-481, 2016.
12. A. V. Dev and A. Mohan,” Recommender system for big data applications based on set similarity of the user preferences” 2016 International Conference of on Next Generation Intelligent Systems (ICNGIS), Kottayam. 2016, pp.1-6.

AUTHORS PROFILE



Abhishek Verma, B. Tech with specialization in Information Technology from SRM Institute of Science and Technology, Kattankulathur, 2020 graduate. Software Engineer. Core Interest: Machine Learning, Algorithm Development, Java, Web Application Development.



V Nallarasana, Assistant Professor (OG) at SRM Institute of Science & Technology, Kattankulathur. M.E with specialization in Computer Science from Karpagam University, 2012. Research Interest: Bigdata, Machine Learning.