iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** II **Month of publication:** February 2021

DOI: https://doi.org/10.22214/ijraset.2021.33043

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue II Feb 2021- Available at www.ijraset.com

Automated Online Course Recommendation System using Collaborative Filtering

Ashuta Thakur[1], Sheetal Rathi[2]
[1, 2]*Department of Computer Engineering, Thakur College of Engineering & Technology, Mumbai, India*

Abstract: ***In the digital world of online information huge amounts of*** **Massive Open Online Courses (MOOCs)** ***are available of different category and domain. Multiple online courses are available on different platform finding appropriate course from this massive available course is difficult for students. Recommender system plays vital role in finding appropriate courses to students. Managing massive amount of information and identifying individual users' choice and behavior has become tedious task nowadays, so the aim of recommender system is to suggest relevant course to student based on user behavior and similarity with another course. Several recommender system techniques are being implemented like content based, collaborative, Knowledge based. This paper aims to build a hybrid approach using collaborative filtering with content base filtering. This system recommendation is based on course description and ratings. Experiments were conducted on real datasets to get the overall performance of proposed system.***

Keywords: Recommendation system, Collaborative Filtering (CF), Content based Filtering (CBF), Knowledge based Filtering, MOOCs.

I. INTRODUCTION

With the ever-growing large volume of online information, recommender systems are an efficient strategy to beat such information overload. Recommender systems are the systems that are designed to recommend things to the user supported various factors. Companies like YouTube, Netflix, Amazon, etc. use recommender systems to help their users to recommend the correct product, video or movies for them. Advances in technology has changed the way of education. Massive Open Online Courses (MOOCs) are capable of providing several learners to access courses over the web. Recommender System (RS) is computerized system that suggest/recommend item to user.

The number of MOOCs and the number of students registered in MOOCs are growing per annum. In 2018, more than 900 universities were offering MOOCs with 11,400 courses available, and around 101 million students had registered in them (Shah, 2018), providing learners with a good sort of choices. With such a high number of courses available, learners now face the matter of choosing courses without being overwhelmed.

With the rise in e-commerce and online business, the number of users interested in online Web services has increased. Both MOOC providers and online businesses advertise their courses and services while learners look for courses that match their interests and needs. In these situations, recommender systems play a crucial role, and have attracted the attention of researchers. Recommender systems are algorithms and techniques that, based on their preferences, suggest matching and related courses or services to the learner, knowledge about which comes from learner profiles and systems-gathered histories. Recommender systems help MOOC providers grow and learners find more appropriate and customized services based on their personalities and interests.

Recommender systems discover patterns in considerable datasets to find out preferences of different users and predict items that correlate to their needs. Recommender systems is divided into two broad categories: collaborative filtering recommender systems and content-based recommender systems. Collaborative filtering recommender systems perform recommendations on users who have had similar taste in the past will make similar choices in the future. Content based recommender systems consider the profile of users and items.

The online course recommendation systems suggest to the students the best courses in which they are interested. This paper presents a recommendation methodology that recommends courses to students based on similarity between courses taken by the target student and other students. It aims to provide an effective course recommendation using multiple techniques. The students will be clustered into groups based on traditional data-mining (DM) techniques will to Collaborative filtering using knowledgebase.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue II Feb 2021- Available at www.ijraset.com

II. RECOMMNDATION TECHNIQUES

Recommendation System is primarily arranged in three sections dependent on the available data. Available data assumes significant part in Recommendation System. Three principal Approaches of Recommendation System are Content based Recommendation System, Collaborative Filtering Recommendation System, and Knowledge Based Recommendation System

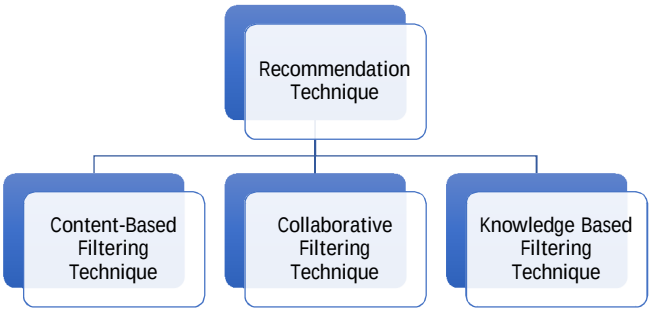


Figure1: Types of Recommendation Technique

A. Content-based Filtering Technique

Content based filtering is based on information filtering examine users' profile in detail for analyzing data. CBF utilize descriptive keywords associated with each item to make recommendations Content based contains information about a user data and choice of user. Mainly, two type of data is required one is users' profile and other is user's interest. The taste is based on how user rates the items. In this CBF technique, rated items will be compare not rated items and similarity between rated items will be evaluated for recommendation. Items that are almost similar to the completely rated ones, will be recommended to the user. Amazon is a best example of content-based filtering. Accuracy of content-based filtering is dependent on user's data. There are two approaches to content-based filtering

1) *Recommendation using User Rating Approach:* In this approach product will be rated on user's choice and similar product rating will be predicted. Advantage of this approach is it works without user reviews.
2) *Recommendation used User Description:* In this type of approach full description if item is required. Detail description of item like title, type, keyword, price, tag line, genre etc is required. This text format strings are converted to number for calculating similarity**.** Term Frequency-Inverse Document Frequency (TF-IDF) is used in extraction features of information.

Drawbacks:

a) Problem of New-user -Same as for collaborative filtering, user profiles are required as input.
b) Limited analysis of content -The recommender highly depends on the information available from the documents. Therefore, the documents contain either some machine-readable text or they need to be classified by users manually. The popularity of two documents, which has the same vectors, cannot be differentiated by such content-based systems.
c) Over specialization-The recommender recommends the items similar to the ones and the one which is already known, leading to a portfolio effect.

B. Collaborative filtering Technique

Collaborative filtering Approach is one of most usually utilized algorithm. This recommender system suggests things dependent on c similarity measures among items and user. The things prescribed to user are liked by comparative User. These comparable user structure a gathering called neighborhood. Assume a user has not evaluated thing suggestion will be given dependent on comparative taste user rating. Collaborative filtering technique has two classification user-based and item based. User based is similitude between the user's choice and item based is comparability among item and different items.

A few downsides of this strategy are as per the following:

1) Problem of New User - For suggesting things, the user should indicate their preferences first without this data, no recommendations can be made.
2) Problem of New-Items – Items that are new in the system need ratings by users before they are being utilized in the recommendation process

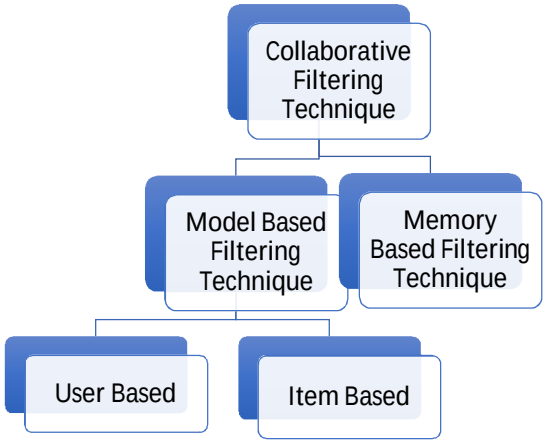


Figure2: Types of Collaborative Filtering Technique

C. Knowledge-based filtering Technique

Knowledge based recommender system depends on domain knowledge and about the user's knowledge. Removing the knowledge of learner's and knowledge about the available information's, is the significant task in knowledge-based recommender system. Knowledge-based recommender systems won't consider assembling long -term generalizations about their users however they favor creating a recommendation supported matching between user's want, preferences and available set of items. This methodology doesn't have the issue of over specialization and sparsity because in this methodology each user is independent with another user and the statistical evidence. Indeed, this methodology is sensitive to change in the learner interest and preferences of learner and furthermore it doesn't have any dependency on information rating. Knowledge-based methodology need not have an initial database of learner's preference and also it is capable to exploit the data concerning the training domain to produce the most effective answer to the learners. However, the point of this methodology is to give the most appropriate recommendations and reasoning about how learning materials of the domain meets the learner's need.

III. RELATED WORK

Many research works have conducted in the field of online course recommender system, content-based filtering, collaborative filtering, data mining and machine learning methods. MOOCs are increasing in huge amount and need for recommender system has also increases already many applications uses recommendation techniques. Applications such as Amazon, Netflix, Facebook, LinkedIn, Twitter etc. uses Recommender system with different approach. We provide a brief overview of the previous research on the development recommendation based on machine learning models in this section.

Amer Al-Badarenah [1] proposed system for course selection based on collaborative filtering technique. Various recommendation system techniques like content-based filtering, classification method, collaborative method, and knowledge-based method where compared. In proposed system recommendation of course is done on basis of similar users. It recommends user course and expected grades for course. Association rule mining technique is applied to discover comparable. The proposed method demonstrated in the experiment demonstrates that association rule mining is a desirable tool for making recommendations. To verify the overall performance of the proposed method, experiments were conducted with real data sets. The downside of the Automated Recommender Method for Course Selection is that it takes a very long time to mine student rules and low performance at the same time.

Sheetal Girase[7], adopted hybrid approach od content-based approach and collaborative filtering approach to recommend colleges to students. College recommending system dataset is collected from 16 IITs, 30 NITs, 20 Government Funded Institutes, and 60 Maharashtra State Engineering Colleges. To decrease the efficiency of algorithm matrix factorization (MF) using Singular Value Decomposition (SVD)is used. System faces big challenge in ranking data which assures user satisfaction. Cold Start is one of the common problems with the recommendation method. As in recommendation system, lot of quality information is needed, but there is situation where there is lack of information. Cold start issue happens when there are new users to the system, new products or a completely new system where there is no data on them, so it is difficult to recommend the recommended system.

Obeidat, R [4] Apriori Algorithm Association Rule Mining and SPADE Algorithm were used in proposed system. Course recommendation is done on three parameter the coverage, the support and the confidence factor of the rule.

Recommending courses for large dataset is difficult as Association rules have high confidence and low coverage on the dataset hence, one low coverage rule will affect the average coverage of the rules.

M. Mohamed [2] The definition of ranking is most common in the recommendation system. Therefore, in the case of a rating-based recommendation system, this data sparsity issue would have this problem, since certain users will score the subset of the overall database because most common products would have less ratings that will impact the accuracy of the recommendation system.

Scalability: The system needs more resources for processing information and forming suggestions for users with the increase in users and objects. The majority of tools available are used to classify users with common interests and similar tastes and descriptions. The combination of various types of available techniques and physical enhancement of structures solves this form of problem. Some computations are also calculated offline to improve the speed of the recommendations online.

Over-specialization: Over-specialization is another problem in the recommendation system, especially in content-based filtering. The primary goal of such systems is to suggest items that are strongly compatible with consumer preferences; however, this could lead to indicating items that have already been visited by the user. For example, in the news recommendation system, which is based on content-based filtering, a user can receive recommended news that she has already read and that the consumer does not like different ones.

Fang Liu [11] Apriori algorithm is introduced in core function design. Designed scheme is applied in one agricultural major category of China Open University system to find student course selection frequent pattern and strong association rules in Animal husbandry and veterinary medicine major category.

User Based Collaborative Filtering (UBCF) approach builds similarity between users by M. Kommineni, P [1]. Recommendation of books using similarity measures proved to be efficient using user based collaborative filtering. Similarity methods were done using PCC, CPCC, Cosine, Jaccard where CPCC (constarined pearson correlation) similarity measure is the best measure for calculating similarities among users.

D. Estrela [6] Proposed hybrid approach by combining Content based filtering and Collaborative filtering approach to recommend online course for elearners. Hybrid approach faces cold start problem and precision of recommender was not accurate.

JIANG Ya-tong [12] A hybrid recommendation system combining the features of collaborative, content based for personalized subject recommendation for students. Each subject score is compared with each student using collaborative filtering. Pearson correlation and angle cosine are used to calculate similarity is proposed system.

Hai-hui Wang [8] Modern learning style MOOCs, massive open online course, provide online courses to unlimited users and open access via the web. It becomes difficult for the students to decide which course to opt within limited time. Jaccard's Similarity is used to cluster data.

V. Manvitha[13]The goal of the proposed system was to solve the issue of cold starts. It is a recommendation system that advises users to play music. The Taste Profile dataset obtained from the Million Song Dataset is the dataset used. 10M is the dataset scale. Moreover, no person relies on online sites to buy songs, apparels, books, rented films, etc. Competition occurred between different online sites that forced owners of the Websites that provide their users with personalized music recommendations. So, the recommender systems came into existence. The issue with cold starting is that the recommenders do not recommend entirely new users or products for those users or items for which adequate knowledge is not available. The combination of association rules and clustering technique is an algorithm based on this music recommender.

Effective Recommender system method was also tried during the research, the most popular would be [14], from Amazon.com. They use recommendation algorithms to personalize the online store for each customer. Rather than matching the user to similar customers, item-to-item collaborative filtering matches each of the user's purchased and rated items to similar items, then combines those similar items into a recommendation list. item-to-item collaborative filtering
algorithm is scalable over very large customer and product catalogs, requires very less processing time to recommend product.

Tanay Kulkarni [3] In proposed system student learning process is being monitor based on available resources so that student's area of interest can been known to recommend. Huge amount of information available on the web so getting relevant information which fulfills user's searching criteria is a main concern. Therefore, to address this issue the primary step of model is to develop a keyword focused web crawler which extracts relevant URLs that includes the keyword searched by the user and considers only such web pages as significant for further processing, thus increasing the accuracy and search efficacy. The system will get URLs for user activity analysis and other users will receive recommendations. The subject of focus in this paper is the use of techniques for content fetching and effective user profiling, so that the system gives out accurate recommendations.

M Viswa Murali [5] Proposed system recommender model "UBCF (User Based Collaborative Filtering) model" generates the recommendation from the closest user behavior using the whole matrix saved. Internally the model calculates the cosine similarity among all consumers represented as vectors.

Kiratijuta Bhumichitr[9]Choosing university courses is complicated, so this method focuses on advising students to take elective courses. Since many elective open courses are available, but students do not have time to explore, this system often allows students to know the correct course to be chosen. It recommends a course based on enrollment data from students.This paper is used by most well-known algorithms to discover comparable students who use Pearson Correlation Coefficient and Alternating Least Square Correlation Coefficient to collaboratively filter (ALS). The experimental results show that with 86 percent accuracy, applying ALS in this domain is superior to collaborative.Student enrollment information was not adequate to assess student conduct to support better recommendation. Downside of this method.

IV. THE PROPOSED RECOMMENDATION SYSTEM

There has recently been an increase in demand for online courses and distance education. Instead of institutes, students tend to study online, so the key explanation for the online course recommendation system is to recommend relevant courses to students. To boast of their skills and profession, students learn from MOOC.They consider factors such as course length, current market patterns for courses, faculty teaching for specific courses, catalogs, interest, etc., to choose the appropriate course.

The proposed system focuses to use collaborative filtering technique with knowledge-based to recommend best appropriate course to students. On available dataset the proposed system will map all similar students based on K means algorithm. After Collaborative filtering technique will be hybrid with knowledge-based to recommend appropriate course to students.

Following is the proposed system flow:

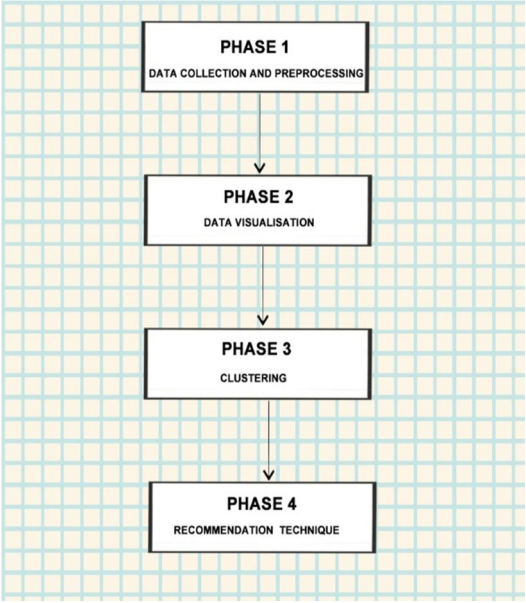


Figure3: Proposed system model flow diagram

1) Step 1: Data Collection and Preprocessing

Data available is very less and insufficient to build a recommender system. Many MOOCs don t provide confidential dataset of their users and courses. This proposed system requires ratings of courses so web scrapping is the technique through which this data was pulled. Web Scraping is a method used to retrieve vast quantities of website data by downloading the data and saving it to a local file on your machine or to a database in table (.csv) format. Data scraping is a effective method to extract data from websites (based on the website s regulations). Dataset of proposed system was of Udemy which is one of best MOOCs platform.

Data available through scrapping is a raw data which needs to be transform in a useful format.Raw Data is cleansed through processes such as filling in missing values or removing incomplete rows of data, smoothing out noisy data, or solving data inconsistencies. Irrelevant data is being dropped from raw data. The cleaned dataset contains live and available public courses.

Name of dataset	df_courses	df_reviews
Number of rows originally	10.000	1.415.734
Number of rows after cleaning	8.834	1.391.194
Final number of columns	30	4

2) Step 2: Data Analysis and Visualization

Data analysis is the heart of the machine learning based projects. After preprocessing of the data next step is to analyze and investigate data sets and summarize their main characteristics by data visualization methods. Data analysis answers the various questions such as what data can tell us before building system, how much data is sufficient? How to establish the relationship between learned data and result. Data analysis help us to actually predict top courses in MOOCs, total subscribers, number of courses which can be helpful for clustering.

Data Analysis helps us to

· to offer insight into a knowledge set.

· Understand the data structure.

· Extract important relationships that hold between various parameter.

It is an honest practice to know the info first and check out to collect as many insights from it.

3) Step 3: Clustering

Aiming to identify similarity in course description and course objective, in this step, course is clustered using the K-means algorithm. K-means clustering is one of the simplest and popular unsupervised machines learning algorithms.

Clustering refers to the task of classifying a set of objects into a set of homogeneous groups such that the objects within the same group (i.e., cluster) are most similar while having the greatest dissimilarity to objects in other groups. The K-means algorithm works by randomly selecting points as the initial centroids of clusters.

A measure of distance (e.g., Euclidean distance) is then calculated for each one of the other points, and each one is assigned to the cluster having the closest centroid. Subsequently, a new centroid is computed for each one of the clusters. This iterative process of assigning points to clusters and updating the centroids continues until the sum of squared errors is minimized. Clusters chosen by proposed system is K=8. Each cluster are label by top 5 words that helps to understand which cluster contain which courses computed for each one of the clusters.

Moreover, PCA is conducted with three, four, and five components for the purpose of dimension reduction.

4) Step 4: Recommendation System

Proposed system tries to focus on Collaborative filtering with knowledge base. Collaborative filtering is based on user preferences whereas knowledge base is dependent on additional knowledge about user from profile and course information.

In proposed system course reviews where not appropriate for collaborative filtering. To resolve this after clustering step additional knowledge of courses features where used with collaborative filtering.

In a knowledge-based algorithm, that includes feature matrix includes:

1) Additional features to recommender system

2) Scaled Feature Matrix

3) Similarity Measure

New features were added to recommender system according to categories in clusters. Different features have different magnitude to scale for example rating of courses varies from 0 to 5 using scikit-learn library features where standardizes. Similarity measures between different courses is done by cosine similarity. Cosine Similarity measures angle between two vectors in multidimension space here two vector refers to addition features of courses. Proposed system recommends course of user's previous course if any or similar course based on current selected course id.

V. RESULTS

We performed experiments using courses dataset taken by Udemy. Top 10 courses records are shown Figure4. Around 900 courses where there without reviews/ratings. The average time period of a course is 26 months. Machine learning course had highest number of subscribers.Clustering algorithm based on the course descriptions showed good results. Multiple k-s number of clusters were tried out in kMeans algorithm. For optimal number of clusters k=8 shown in figure5.MAEcalculates the average magnitude of the errors without considering their path.

$$\mathrm{MAE} = \frac{1}{|\hat{R}|} \sum_{\hat{r}_{ui} \in \hat{R}} |r_{ui} - \hat{r}_{ui}|$$

r^ui: ratings predicted

rui: ratings known

The deviation between predicted rating and the actual rating of different users and items are measured using the above formula. Lower MAE values indicate higher accuracy. Overall prediction accuracy of item based and user-based algorithms is very high as MAE values for all possible neighborhood size falls in the range of 0.33 to 0.38. From the observation of figure 6 item based and user-based algorithm results, there is not much difference in accuracy but there is lots of difference in proposed algorithm.

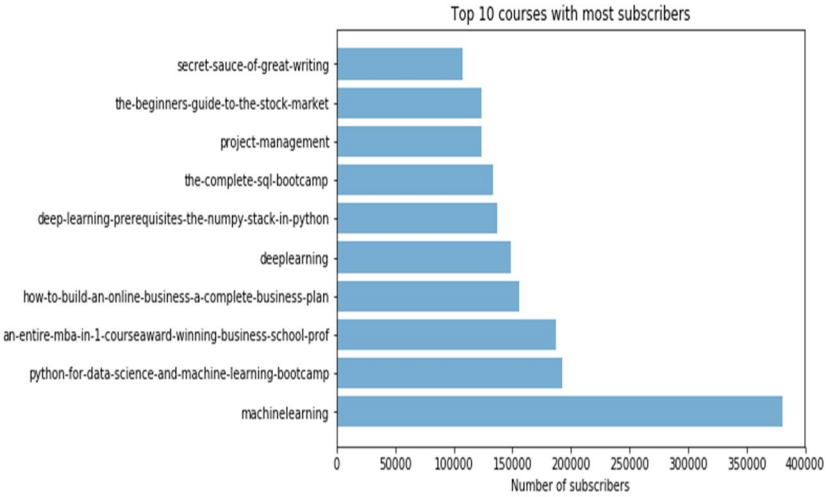


Figure4: Top 10 courses

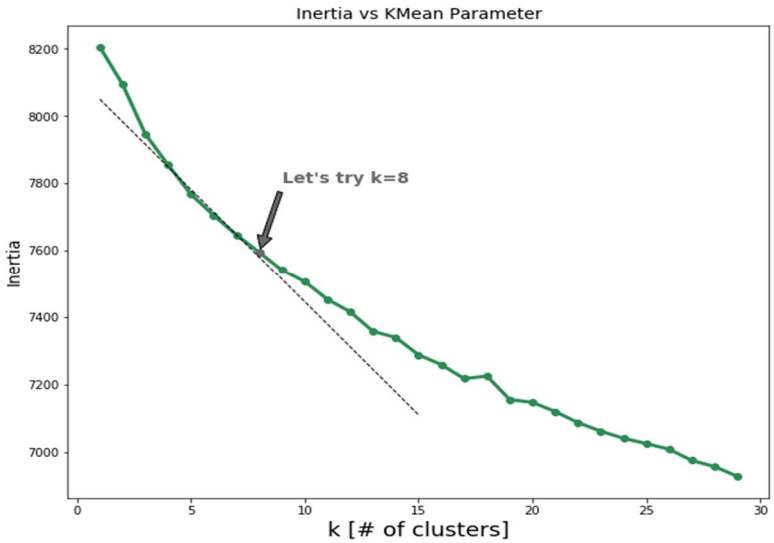


Figure5: kMean parameter

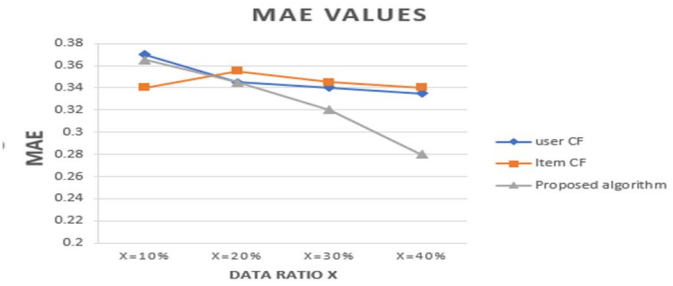


Figure6: MAE

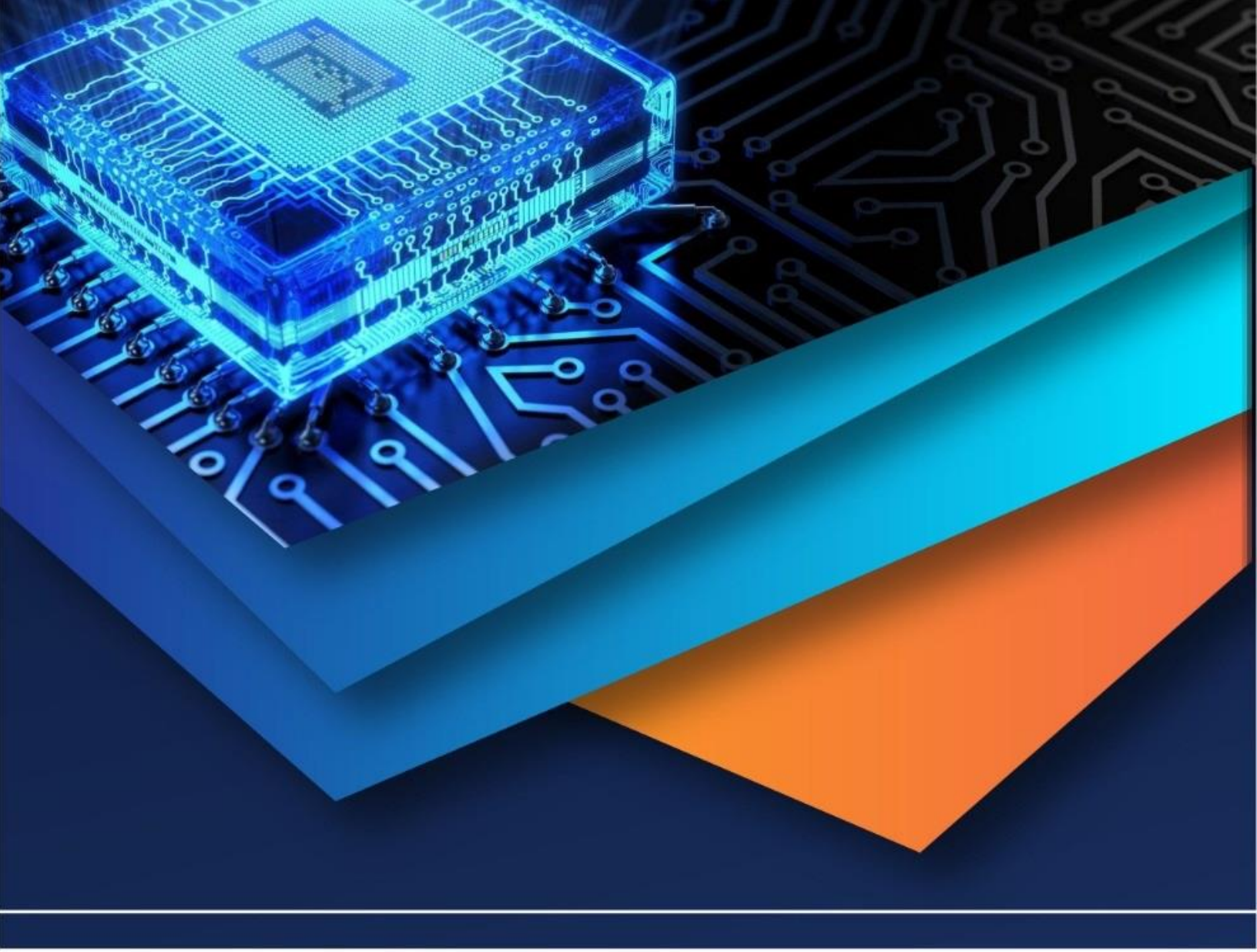
VI. CONCLUSION

Course recommendation system help the students in course choosing process. Classical recommendation techniques like content-based filtering, collaborative filtering, and knowledge-based filtering are used in different commercial websites. For recommending course to students' similarity is performed for course topics and keywords. Proposed system recommends course to students based on users rating and course description. This online course recommender system focused on hybrid approach of collaborative and content-based filtering to recommend courses to students. The similarity measures between the user and courses are precise by making use of the cosine similarity technique. The user is suggested with the top-rated courses according to his interest.

In the future, we can use other information of user so that it would determine the behavior of the student using NLP for further recommendation.

REFERENCES

[1] Madhuri Kommineni , P.Alekhya, T. Mohana Vyshnavi, V.Aparna, K Swetha, V Mounika,“ Machine Learning based Efficient Recommendation System for Book Selection using User based Collaborative Filtering Algorithm” IEEEAugust 2020.
[2] M. Mohamed, M. Khafagy and M. Ibrahim, "Recommender Systems Challenges and Solutions Survey", 2019 International Conference on Innovative Trends in Computer Engineering (ITCE), 2019.
[3] Tanay Kulkarni, Madhur Kabra , Dr. Radha Shankarmani ,“ User Profiling Based Recommendation System For E-Learning” IEEE2019.
[4] Obeidat, R., Duwairi, R., Al-Aiad, A. ,“ A Collaborative Recommendation System for Online Courses Recommendations ” IEEE2019.
[5] M Viswa Murali, Vishnu T G, Nancy Victor ,” A Collaborative Filtering based Recommender System for Suggesting New Trends in Any Domain of Research ” International Conference on Advanced Computing & Communication Systems (ICACCS) 2019.
[6] Estrela, D., Batista, S., MartinhoD., & Marreiros, G., “A Recommendation System for Online Courses”, Springer International Publishing AG 2017.
[7] Sheetal Girase, Varsha Powar and Debajyotimukhopadhyay, “A User-friendly College Recommending System Using User-profiling And Matrix Factorization Technique”,International Conference on Computing, Communication and Automation, 10.1109/CCAA.2017.8229779.
[8] Hai-hui Wang, Chalothon Chootong, Ankhtuya Ochirbat, Worapot Sommool, W K T M Gunarathn, and TimothyK.Shih, “Online Courses Recommendation System Based on Industry Occupation Skills Requirements”, IEEE, 2017.
[9] Kiratijuta Bhumichitr, Songsak Channarukul, Nattachai Saejiem, Rachsuda Jiamthapthaksin and Kwankamol Nongpong, “Recommender Systems for University Elective Course Recommendation”, IEEE, 2017.
[10] Amer Al-Badarenah and Jamal Alsakran, “An Automated Recommender System for Course Selection”, International Journal of Advanced Computer Science and Applications, vol.07, pp. 3, 2016.
[11] Fang Liu, Shaogang Zhang, Junyi Ge, Feng Lu and Jun Zou, “Agricultural Major Courses Recommendation Using Apriori Algorithm Applied in China Open University System”, International Symposium on Computational Intelligence and Design, 10.1109/ISCID.2016.1109.
[12] JIANG Ya-tong, fu Qiang, li Fei, lv Hai-xia, wu Gang, and Mehmet Fatih Usl, “Personalized Recommendation and Analysis Method for Student Partiality for One or Some Subject(s) In HigherEducationManagement”, IEEE, 10.1109/ICMSE.2014.6930479.
[13] V. Manvitha, M. Sunitha Reddy “Music Recommendation System Using Association Rule Mining and Clustering Technique to Address Cold start Problem”, International journal of engineering and computer science,Vol. 3, No. 6855-6858, 2014.
[14] Linden G, Smith B, York J. Amazon.com recommendations: Item- to- item collaborative filtering[J]. IEEE Internet Computing, 2003, 7(1): 2003: 76- 80.
[15] Xiwang Yang, Chao Liang, and Miao Zhao, “Collaborative Filtering-Based Recommendation of Online Social Voting”, IEEE Transactions on Computational Social Systems,10.1109/TCSS.2017.2665122.
[16] Dhruv Shah, Pratik Shah and Asim Banerjee, “Similarity Based Regularization for Online Matrix-factorization Problem: An Application to Course Recommender Systems”, IEEE, 10.1109/TENCON.2017.8228164.
[17] Suleyman Uslu, Can Ozturan, and Mehmet Fatih Usl, “Course Scheduler and Recommendation System for Students”, IEEE, 10.1109/ICAICT.2016.7991812.
[18] Skills Requirements”, IEEE, 10.1109/UMEDIA.2017.8074083.
[19] Kaustubh Kulkarni, Keshav Wagh, Swapnil Badgujar and Jijnasa Patil, “A Study of Recommender Systems with Hybrid Collaborative Filtering”,International Research Journal of Engineering and Technology (IRJET), Volume: 03 ,2016.
[20] Yiu-kai Ng and Jane Linn, “Crsrecs: A Personalized Course Recommendation System for College Students”, International Journal of Advanced Computer Science and Applications, 10.1109/IISA.2017.8316368.
[21] S. Ray and A. Sharma, "A Collaborative Filtering Based Approach for Recommending Elective Courses", Information Intelligence, Systems, Technology and Management, pp. 330-339, 2011.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY