Jagannath University Research Journal (JURJ) Volume No.-V - , Issue No. I, Month, April, Year, 2024 ISSN: 2582-6263

Smart Recommender System for Online Courses

Kalpana [a], Dr. Renu Bagoria[b] and Dr. Praveen Arora[c]
[a]Jagannath University, Jaipur India,[b] Jagannath University,Jaipur India,[c]JIMS Rohini Delhi, India

Abstract: This research paper investigates recommender systems for online courses offered on Coursera. We explored two recommendation approaches: popularity-based and content-based. The popularity-based approach recommends courses with high ratings and a large number of enrollments. The content-based approach recommends courses with similar skillsets to a course the user has already shown interest in. In our model, Data Privacy Concerns is taken into consideration that develop recommender systems that ensure user privacy by anonymizing data and providing users control over the information used for recommendations. This is ensured that the recommendation is made on the basis of user choice. It is very important to convince the Learner that the recommended course is the best suited course as per the choice made by the learner. The main focus of the paper is to control with the course dropout rate as a large number of leaners enroll in a course but dropout from the course due to several reasons like lack of interest and other issues.

Keywords: Recommender System, Popularity Based Recommender, Content Based Recommendation, MOOCs

I. INTRODUCTION

Massive Open Online Courses (MOOCs) like Coursera have revolutionized access to education, offering a vast array of courses to learners worldwide. However, with a growing course catalogue, navigating course options and identifying the most relevant ones can be a challenge for users. Recommender systems play a crucial role in addressing this challenge by suggesting courses that align with a user's interests and learning goals.[2]

This research investigates the development of recommender systems for Coursera courses. We explore two primary approaches: popularity-based and content-based. The popularity-based approach prioritizes courses with high ratings, a large number of enrolments, and potentially recent trending topics. The content-based approach focuses on recommending courses that share similar skillsets with a course the user has already shown interest in.

This paper leverages a dataset containing information about Coursera courses, including titles, ratings, skills covered, and enrolment numbers. We employ data cleaning and pre-processing techniques to prepare the data for analysis. Subsequently, we implement both popularity-based and content-based recommender system algorithms. [3,4]

The popularity-based system considers course ratings, enrolment figures, and potentially factors like recent enrolment trends to recommend highly-rated and well-populated courses. The content-based system analyses course descriptions and skill tags to identify courses that share similar knowledge domains with a user's chosen course of interest.

1.1 Objectives

This research contributes to the field of recommender systems for e-learning platforms by:

- Implementing and evaluating both popularity-based and content-based recommender systems for Coursera courses.
- Proposing a weighted recommendation score that combines insights from both approaches.
- Highlighting the potential of recent advancements in deep learning and user modeling for future research directions.

II. LITERATURE SURVEY

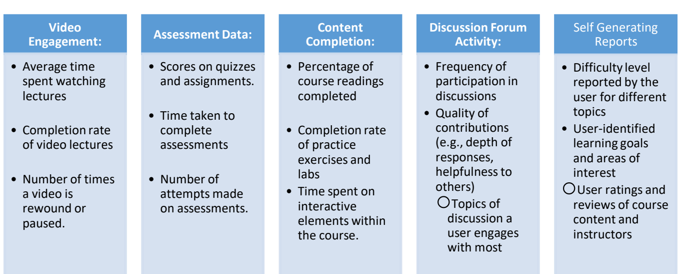
Online course recommender systems can provide more personalized and effective learning experiences.[1] They can recommend courses that align with a user's Learning pace that is the Data on video engagement and content completion can identify users who prefer shorter, concise content or those who benefit from deeper dives into specific topics. Learning goals: Self-reported data and forum participation can highlight a user's aspirations [9]. Strengths and weaknesses: Assessment data and

Jagannath University Research Journal (JURJ)

Volume No.-V - , Issue No. I, Month, April, Year, 2024 ISSN: 2582-6263

practice activity can reveal areas where a user excels and topics that require further reinforcement. Learning style preferences (visual vs. auditory) might be inferred based on video engagement patterns.[10]. Research on recommender systems that utilize course descriptions, skill tags, or learning objectives of content-based recommendations. Papers explore techniques for extracting relevant keywords or skills from course descriptions would be relevant here [8]. Taking into consideration the popularity of specific type of course, research on recommender systems that leverage user ratings, course reviews, and enrollment numbers for course recommendations. Studies that explore optimal weighting schemes for these factors would be valuable. A lot of research is being done in the direction of Evaluation Metrics for Recommender Systems that discusses various metrics used to evaluate the effectiveness of recommender systems in e-learning settings. This could include click-through rates (CTR), conversion rates (course enrollment), and user satisfaction surveys.

Figure 1. User Performance Matrix



Learning analytics data that could be used by a recommender system to personalize online course recommendations. By incorporating this type of learning analytics data, a recommender system can gain a more nuanced understanding of a user's learning style, pace, and areas of strength and weakness.[6,7]

III. Proposed Recommendation Model

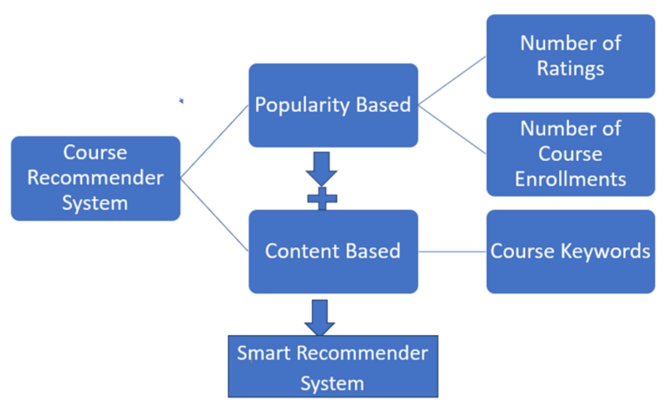
We propose a hybrid recommendation model that incorporates elements of both popularity and content-based approaches for a more comprehensive evaluation. [11,12] Here recommendations can be made using a Smart Recommender System that will provide recommendations on the basis of the following important Concerns:

Data Privacy: In order to ensure the data privacy, the user is allowed to decide that whether he wants to disclose his personal details while providing review/rating about the course.

Cold Start Problem: To address the challenge of recommending new courses with limited data on user interactions or course content. This might involve leveraging transfer learning from similar courses or using collaborative filtering techniques more effectively.[14]

Reasoning to users. This could involve highlighting the specific skills or learning objectives that connect a recommended course to the user's interests.

Figure 2. Proposed Model



Dataset For our work, we have taken the Coursera Dataset having 1000 entries and 12 attributes. The different attributes of the dataset is as follows in the figure shown below

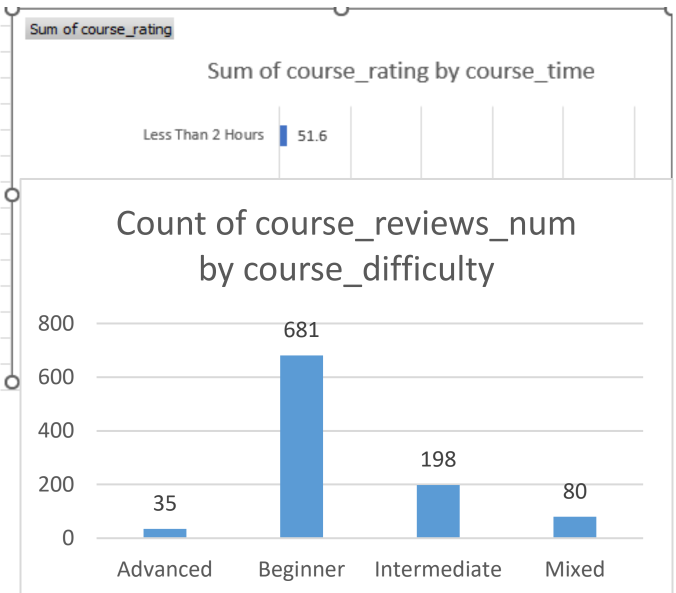
Figure 3. Attribute list of Dataset

```
Data columns (total 12 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   course_title              1000 non-null   object
 1   course_organization       1000 non-null   object
 2   course_certificate_type   1000 non-null   object
 3   course_time               1000 non-null   object
 4   course_rating             994 non-null    float6
 5   course_reviews_num        994 non-null    object
 6   course_difficulty         1000 non-null   object
 7   course_url                1000 non-null   object
 8   course_students_enrolled  959 non-null    object
 9   course_skills             1000 non-null   object
 10  course_summary            1000 non-null   object
```

JURJ
Jagannath University Research Journal (JURJ)

Jagannath University
Volume No.-V - , Issue No. I, Month, April, Year, 2024 ISSN: 2582-6263

Figure 4. Course Rating and Course Duration



The dataset used for this research is a CSV file named "coursera_courses.csv". It is assumed that this file contains information about various courses offered on Coursera, including titles, ratings, reviews, skills covered, and enrollment numbers.

Popularity based Recommender System

Figure 6. Number of Reviews on the basis of Course level

	course_title	num_ratings	course_rating	course_url	course_skills	course_students_enrolled
0	(ISC)² Systems Security Certified Practitioner...	492	4.7	https://www.coursera.org/specializations/sscp-...	['Risk Management', 'Access Control', 'Asset',...	6,958
1	.NET FullStack Developer	51	4.3	https://www.coursera.org/specializations/dot-n...	['Web API', 'Web Development', 'Cascading Styl...	2,531
2	21st Century Energy Transition: how do we make...	62	4.8	https://www.coursera.org/learn/21st-century-en...	[]	4,377
	A Crash				['Instrumental	

1. **Data Cleaning and Preprocessing**: We first clean the data by handling missing values and duplicates.
2. **Feature Engineering**: We create a new dataframe,num_rating_df,containing relevant features for the popularity-based recommender system, including course title, number of reviews, average rating, and number of enrolled students.

Figure 7. Average Rating of Courses

	course_title	avg_ratings
0	(ISC)² Systems Security Certified Practitioner...	4.7
1	.NET FullStack Developer	4.3
2	21st Century Energy Transition: how do we make...	4.8
3	A Crash Course in Causality: Inferring Causal...	4.7
4	A life with ADHD	NaN
...	...	...
988	Étudier en France: French Intermediate course ...	4.8
989	Цифровий маркетинг і електронна комерція від G...	4.9
990	تحليلات البيانات من Google	4.8
991	用 Python 做商管程式設計（一）(Programming for Business C...	4.9
992	用 Python 做商管程式設計（二）(Programming for Business C...	4.6

3. **Merging Dataframes:** We merge two dataframes, num_rating_df and avg_rating_df (containing average ratings for each course title), to get a combined dataframe with both course-level popularity metrics (number of ratings and average rating).

Figure 8. Number of Rating and Course Ratings

Out[9]:

	course_title	num_ratings	course_rating	course_url	course_skills	course_students_enrolled
0	(ISC)² Systems Security Certified Practitioner...	492	4,7	https://www,coursera,org/specializations/sscp-...	['Risk Management', 'Access Control', 'Asset',...	6,958
1	.NET FullStack Developer	51	4.3	https://www.coursera.org/specializations/dot-n...	['Web API', 'Web Development', 'Cascading Styl...	2,531
2	21st Century Energy Transition: how do we make...	62	4.8	https://www,coursera,org/learn/21st-century-en...	[]	4,377
	A Crash				['Instrumental	

Jagannath University Research Journal (JURJ)

Volume No.-V - , Issue No. I, Month, April, Year, 2024 ISSN: 2582-6263

4. **Filtering and Recommendation:** We filter courses with a minimum number of ratings (set to 250 in this example) and sort them by their average rating in descending order. This provides a list of popular courses based on a combination of the number of enrolled students and the average rating they received.

Figure 9. Popularity Based Recommendation

```
In [13]: popular_df[popular_df['num_ratings']>=250].sort_values('avg_ratings',ascending=False).head(50)
Out[13]:
```

	course_title	num_ratings	course_rating	course_url	course_skills	course_students_enr
998	用 Python 做商管程式設計 (一) (Programming for Business C...	814,0	4,9	https://www,coursera,org/learn/pbc1	[]	3
943	The Science of Well-Being for Teens	360,0	4,9	https://www.coursera.org/learn/the-science-of-...	[]	11
438	Google Professional Workspace Administrator	563,0	4,9	https://www,coursera,org/professional-certific...	[]	1
391	Fundraising and Development	463,0	4,9	https://www,coursera,org/specializations/fundr...	['Fundraising']	2

Content-Based Recommender System

1. **Skill Extraction:** We extract unique skills from the "course_skills" column, which presumably contains a list of skills covered in each course**.**
2. **Skill Frequency:** We calculate the frequency of each skill **across** all courses to identify the most common skills offered on Coursera.
3. **User Modeling:** We create a user model based on a course the user has shown interest in. This is **achieved** by extracting the skills associated with that particular course.
4. **Course-Skill Matching:** We calculate a similarity score between the user's course (skills) and other **courses** based on the number of skills they have in common

Figure 10. Course Details based on Skills

, 1:00 PM Course Recommender System - Jupyter Notebook

```
In [15]: courses_skills = num_rating_df.copy()
         all_skills = [skill for skills_list in courses_skills['course_skills'] for skill in eval(skills_list)]
         skill_counts = Counter(all_skills)
         skills_df = pd.DataFrame(list(skill_counts.items()), columns=['skill', 'count'])
         skills_df = skills_df.sort_values(by='count', ascending=False)
         print(skills_df)

                                skill  count
164                Python Programming     68
84                      Data Analysis     63
27                   Machine Learning     50
85        Data Visualization (DataViz)     43
35                       Data Science     40
...                               ...    ...
1037  Generative Adversarial Networks      1
1036                    Discriminator      1
1035    glossary of computer graphics      1
1034       Image-to-Image Translation      1
2303                   برنامج Tableau      1

[2304 rows x 2 columns]
```

5. **Weighted Recommendation**: We create a weighted **recommendation** score that considers the similarity score, course popularity (number of ratings), and average rating of each course.
6. **Recommendation:** We recommend courses with the highest weighted scores, excluding the course the user has already shown interest in. Additionally, we identify the **common** skills between the recommended courses and the user's course of interest. Popularity Based Recommendation.

Figure 11. Course Based Recommendation

	course_title	num_ratings	course_rating	course_url	course_skills	course_students_enrolled	co
302	European Business Law	967.0	4.8	https://www.coursera.org/specializations/europ...	[Compete on the Internal Market, Legal Researc...	18,683	3
90	Automate Cybersecurity Tasks with Python	915.0	4.7	https://www.coursera.org/learn/automate-cybers...	[Computer Programming, Python Programming, Cod...	68,630	1
84	Astronomy: Exploring Time and Space	909,0	4,8	https://www,coursera,org/learn/astro	[Solar Systems, Chemistry, Theory Of Relativit...	66,581	1
119	Business Data Management and Communication	890,0	4,6	https://www,coursera,org/specializations/busin...	[Asset Management, Business Analytics, Financi...	11,336	3
669	Make the Sale: Build, Launch, and Manage E-com...	878.0	4.8	https://www.coursera.org/learn/make-the-sale	[Fulfillment and delivery, Website Structure, ...	81,460	1

IV. CONCLUSION

This research paper explores two recommender system approaches for Coursera courses: popularity-based and content-based. The popularity-based approach recommends courses based on their overall ratings and enrollment numbers, while the content-based approach personalizes recommendations based on a user's course interests and the skills they aim to develop. Both approaches can be valuable for suggesting relevant courses to Coursera users keeping in mind the privacy of the user.

V. Future Scope

The field of recommender systems for online courses like Coursera is constantly evolving[15]. Here are some exciting possibilities for future exploration:

- Deep Learning and Neural Networks: Explore the use of deep learning architectures like recurrent neural networks (RNNs) or convolutional neural networks (CNNs) to analyze course content, user interactions, and learning pathways. This could enable more nuanced understanding of course relationships and user preferences.[8]

- Knowledge Graphs and Embeddings: Utilize knowledge graphs to represent relationships between courses, skills, and career paths. Embed these entities in a low-dimensional vector space to capture latent relationships and enable more sophisticated recommendation algorithms.

REFERENCES

[1]. Kalpana, Chauhan, S.S., Singh, M.K., Bagoria, R.," Novel Taxonomy for E-learning Recommender System Using Opinion Mining", In: Garg, D., Narayana, V.A., Suganthan, P.N., Anguera, J., Koppula, V.K., Gupta, S.K. (eds) Advanced Computing,2022.

[2]. Jalota, Chietra & Agrawal, Rashmi," Analysis of Educational Data Mining using Classification",pp 243-247,2019

[3]. Uddin, A. S. Imran, K. Muhammad, N. Fayyaz and M. Sajjad, "A Systematic Mapping Review on MOOC Recommender Systems," in *IEEE Access*, vol. 9, pp. 118379-118405, 2021, doi: 10.1109/ACCESS.2021.3101039.

[4]. Yoo, J., Yoo, S., Lance, C., Hankins, J.: Student Progress Monitoring Tool Using Treeview. In: The 37th Technical Symposium on Computer Science Education, SIGCSE'06. ACM Press. March 1-5, Houston, USA (2006) 373-377

[5]. Papia Bawa1,Retention in Online Courses: Exploring Issues and Solutions—A Literature Review, SAGE Open January-March 2016: 1–11© The Author(s) 2016.

[6]. Stanford-Bowers, D. E. (2008). Persistence in online classes: A study of perceptions among community college stakeholders. MERLOT Journal of Online Learning and Teaching, 4, 37-50.

[7]. Tariq, Mustafa & Habib, Hafiz. (2024). A Reinforcement Learning Based Recommendation System to Improve Performance of Students in Outcome Based Education Model. IEEE Access. PP. 1-1. 10.1109/ACCESS.2024.3370852.

[8]. Deepak, Gerard & Trivedi, Ishdutt. (2023). A Hybridized Deep Learning Strategy for Course Recommendation. International Journal of Adult Education and Technology. 14. 1-16. 10.4018/IJAET.321752.

[9]. Bower, M.S., Torrington, J., Lai, J.W., Petocz, P., & Alfano, M. (2024). How should we change teaching and assessment in response to increasingly powerful generative Artificial Intelligence? Outcomes of the ChatGPT teacher survey. Education and Information Technologies.

[10]. Park, Dong Hyun, et al. "A hybrid recommendation system based on collaborative filtering and content-based filtering for e-learning systems." Expert Systems with Applications 39.3 (2012): 3363-3370.

[11]. Baldiris, Konstantinos, et al. "A hybrid recommender system for e-learning platforms." Journal of Systems and Information Technology 16.1 (2014): 1-13.

[12]. Parra-Diaz, William, and Jaime Andres Gutierrez-Garcia. "A content-based recommender system for adaptive learning environments." IEEE Transactions on Learning Technologies 8.4 (2015): 302-314.

[13]. Delgado, Javier, et al. "Building a hybrid recommender system for a MOOC platform." Knowledge-Based Systems 116 (2017): 244-257.

[14]. McNee, Sean M., et al. "On the recommender system cold start problem." ACM Transactions on Information Systems (TOIS) 20.1 (2002): 16-26

[15]. Parra-Diaz, William, and Jaime Andres Gutierrez-Garcia. "A content-based recommender system for adaptive learning environments." IEEE Transactions on Learning Technologies 8.4 (2015): 302-314.