# DEVELOPING A SMART COLLABORATIVE FILTERING SYSTEM IN THE DEVELOPMENT OF AN EFFICACIOUS COURSE RECOMMENDATION SYSTEM

**Arnav Chawla**

*Bharat Mata Saraswati Bal Mandir, Narela, New Delhi*

## ABSTRACT

*In the era of the digital world, huge measures of Massive Open Online Courses (MOOCs) are available in various classifications and spaces. Various web-based courses are accessible at various stages. Tracking down a proper course from this huge available course is hard for students. The recommender system is crucial in tracking down appropriate courses for students.*

*Dealing with the enormous measure of data and distinguishing individual clients' decisions and conduct has become monotonous. Consequently, the recommender system proposes the relevant courses to understudy given client behaviour and comparability with another course. This system suggestion depends on course descriptions and reviews. A few recommender framework methods are used: content, collaborative, and Knowledge-based. This paper is designed to construct a hybrid approach using cooperative and content-based extraction. Tests were directed on real datasets to get them by and large execution of the proposed framework.*

## INTRODUCTION

With the steadily developing huge volume of online data, recommender frameworks are proficiently defeating such data overload. Recommender systems are intended to prescribe things to the client, supported by different variables. Propels in innovation have significantly altered the method of training. MOOCs can give a few students admittance to online courses. Recommender System (RS) is an electronic system that proposes/prescribes things to the client. Organizations like YouTube, Netflix, Amazon, and so forth use recommender frameworks to assist their clients with suggesting the right article, video or film. The quantity of MOOCs and students enlisted in MOOCs are developing per annum. In 2018, more than 900 colleges were offering MOOCs with 11,400 courses accessible, and around 101 million students had enrolled for them (Shah, 2018), making great decisions. With such countless courses accessible, students currently face the question of picking courses without being impressed.

With the ascent in web-based and online business, the quantity of clients willing on internet-based Web administrations has grown. MOOC suppliers and online organizations promote their courses and administrations while students search for courses that match their inclinations and needs. In these circumstances, the recommender system assumes an urgent part and stands out among scientists. Recommender systems are analyses and procedures that recommend related courses or administrations to the student, given their inclinations. Recommender frameworks help MOOC suppliers develop, and students see additionally suitable and renovated management given their characters and interests. Information comes from student profiles and frameworks accumulated accounts.

Recommender frameworks find designs in enormous datasets to decide the preferences of various clients and anticipate things that associate with their requirements. Recommender frameworks are split into two general classes: cooperative sifting recommender frameworks

and happy-based recommender frameworks. Cooperative sifting recommender frameworks perform suggestions on clients with comparable preferences and will pursue comparable decisions later on. Content-based recommender frameworks think about the profile of clients and things.

The web-based course recommendation system recommends the best courses they are keen on to the understudies. This paper presents a proposal procedure that prescribes courses to understudies given similitudes between courses taken by the objective students and researchers. It means to give a compelling course proposal utilizing various procedures. The understudies will be clustered into bunches in light of traditional data mining (DM) procedures will to Collaborative sifting utilizing knowledgebase.

## THE PROPOSED RECOMMENDATION SYSTEM

There has, as of late, been an expansion popular for online courses and distance schooling. Rather than establishments, understudies will generally study on the web, so the vital clarification for the internet-based course suggestion framework is to prescribe significant courses to understudies. To flaunt their abilities and calling, understudies gain from MOOCs. They consider factors, for example, course length, current market designs for courses, personnel education for explicit courses, inventories, interests, and so on, to pick the suitable course.

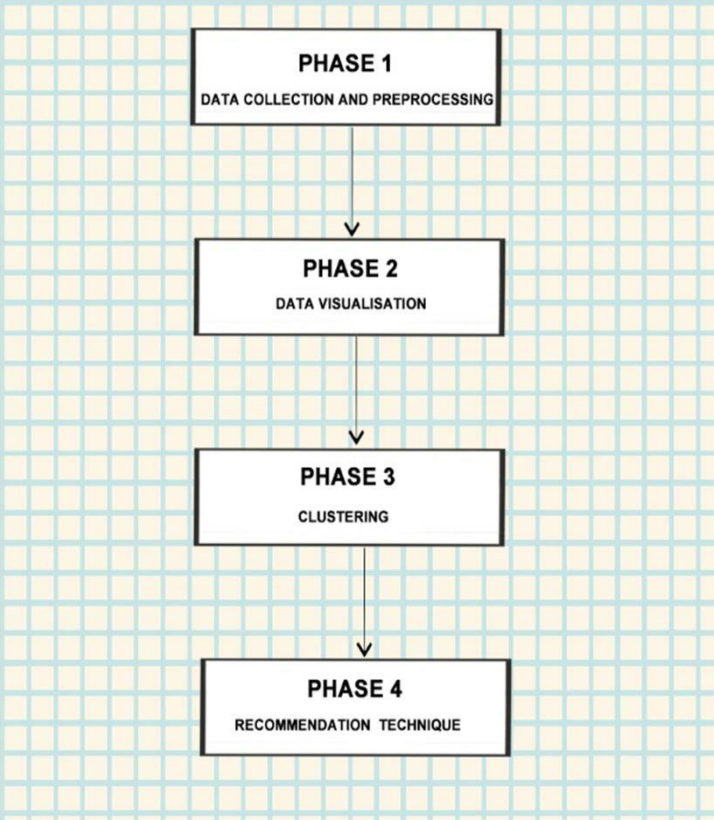


Fig 1: Flow diagram of Proposed System

The proposed framework utilizes information-based cooperative separating procedures to prescribe the best suitable course to understudies. The proposed framework will plan all comparable understudies given the K means calculation on the accessible dataset. After the Collaborative sifting method will be half and half with information based to prescribe the fitting course to understudies.

**1) Step 1: Data Collection and Pre-processing**

Information accessible is extremely less and inadequate to fabricate a recommender framework. Numerous MOOCs wear t give classified datasets of their clients and courses. This proposed framework requires evaluations of courses, so web rejecting is the strategy through which pulled this information. Web Scraping is a technique to recover huge amounts of site information by downloading it and saving it to a nearby document on your machine or a data set in table (.csv) design. Information scratching is a powerful technique to extricate information from sites (based on the site s guidelines). The dataset of the proposed framework was from Udemy, one of the most outstanding MOOCs stages.

Crude Data is purged through cycles like filling in missing qualities, eliminating short information lines, streamlining uproarious information, or addressing information irregularities. The cleaned dataset contains live and accessible public courses. Unimportant information is being dropped from raw data. Information accessible through rejecting is basic information which should change into a valuable configuration.

**2) Step 2: Data Analysis and Visualization**

Data analysis is the heart of machine learning-based projects. After pre-processing the data, the next step is to analyse and investigate data sets and summarize their main characteristics using data visualization methods. Data analysis answers various questions, such as what data can tell us before building a system and how much data is sufficient. How to establish the relationship

between learned data and results. Data analysis helps us to predict top courses in MOOCs, total subscribers, and the 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 parameters.

Knowing the info first and checking it out is ethical to collect as many insights from it.

**3) Step 3: Clustering**

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

Clustering refers to 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 randomly selects points as the initial centroids of clusters.

A measure of distance (e.g., Euclidean distance) is then calculated for each of the other points, and each one is assigned to the cluster with 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. The cluster chosen by the proposed system is K=8. Each cluster is labelled with the top 5 words that help to understand which cluster contains which courses are computed for each cluster.

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

**4) Step 4: Recommendation System**

The proposed system tries to focus on Collaborative filtering with the knowledge base. Collaborative filtering is based on user preferences. The knowledge base is dependent on additional knowledge about a user from profile and course information.

In the proposed system, course reviews were not appropriate for collaborative filtering. After the clustering step, additional knowledge of course features was used with collaborative filtering.

A knowledge-based algorithm that includes a feature matrix includes:

1) Additional features to the recommender system

2) Scaled Feature Matrix

3) Similarity Measure

New features were added to the 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 standardized scikit-learn library features. Similarity measures between different courses are done by cosine similarity. Cosine Similarity measures the angle between two vectors in multidimensional space. Here, two vectors refer to additional features of courses. The proposed system recommends courses of the user's previous course, if any, or similar courses based on the currently selected course id.

## RESULTS

We performed experiments using the course dataset taken by Udemy. The top 10 course records are shown in Figure4. Around 900 courses were there without reviews/ratings. The average period of a course is 26 months. The machine learning course had the highest number of subscribers. The clustering algorithm based on the course descriptions showed good results. Multiple k-s numbers of clusters were tried out in the K-means algorithm. For an optimal number of clusters, k=8 is 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 the predicted rating and the actual rating of different users and items is measured using the above formula. Lower MAE values indicate higher accuracy. Overall prediction accuracy of an item based and user-based algorithms are very high as MAE values for all possible neighbourhood 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 a lot of difference in the proposed algorithm.

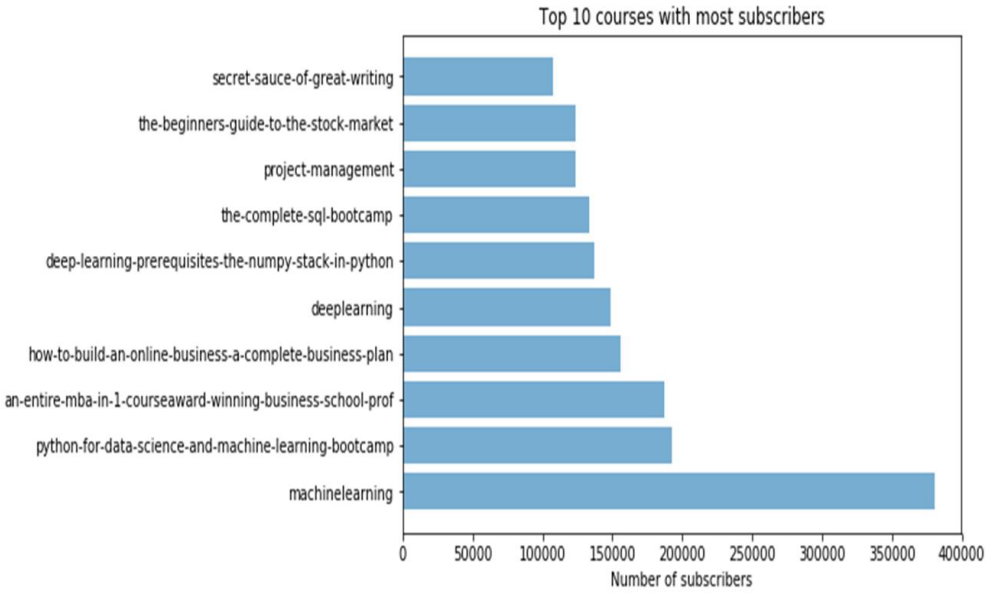


Fig 2: Top 10 Courses


Fig 3: MAE

## CONCLUSION

Course recommendation systems help the students in the course choosing process. Classical recommendation techniques like content-based filtering, collaborative filtering, and knowledge-based filtering are used in different commercial websites. For recommending courses to students' similarity is

performed for course topics and keywords. The proposed system recommends courses to students based on users rating and course descriptions. This online course recommender system focused on a hybrid approach of collaborative and content-based filtering to recommend courses to students. Using the cosine similarity technique, the similarity measures between the user and courses are precise. The user has suggested the top-rated courses according to his interest.

## 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" IEEE August 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" IEEE 2019.

[4] Obeidat, R., Duwairi, R., Al-Aiad, A., "A Collaborative Recommendation System for Online Courses Recommendations" IEEE 2019.

[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 Debajyoti Mukhopadhyay, "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.