Course Recommendation System for E-Learning

Pakshal Shah, Shrikrishna Munde, Azaan Farooq, Giriprasad Hanamsagar
Prof. Reena Pagare

A Department of Computer Science and Engineering,
MIT School of Engineering, MIT-ADT University, Pune, 412201, India

Abstract:- Since the pandemic of 2020, Online learning platforms have boomed. People worldwide are eager to learn and explore new things during their free time. Also, with this emerging world of technologies, contents are available online for free but present in huge chunks. This freely available content is cluttered, due to which many people face problems in reaching their goals with proper guidance free of cost. Speaking of E-Learning, it is mainly focused on two students, i.e., a) Traditional people who only gain knowledge before jumping to professional life, and the other student is b) Workers who have prior knowledge but need to improve what they already have to keep up with the job requirement, this learning can be called lifelong learning. There are several resources for learning as a result of technological innovation. different sources for learning, i.e., For notes, there are many resources and learning from video sources there are countless numbers of video content which leads users to churn from one source to another and thus leads to distraction from goal without any proper guidance. To overcome this, we build a course recommendation system that will help learners choose the right a ?.

Keywords:- *E-Learning, Online Course Recommendation System.*

I. INTRODUCTION

The improveme**1 isw1a2kp kn**nt of recommendation systems for online learning has been the subject of extensive research; with this project, we aim to research and test different Machine learning algorithms and their combinations to achieve good results. But first, it's important to understand why recommendation systems were developed. In today's world, where the internet is filled with a huge amount of information, it is very hard to classify and decide which information is more suitable to the context or more accurate, so to solve this, recommendations were designed to filter out the extra clutter and produce the desired or similar results. In the case of E-learning, the internet is filled with a lot of resources to learn from, But it's hard to choose the right material or course which can help them in achieving their goals. Even the websites that promise to help in such cases often end up bombarding learners with the course suggestions, which again causes cognitive overload, and It becomes hard to decide what to choose and where to learn from. We approached this problem by studying different recommendation algorithms and their combinations, such as the Random Prediction Method, Popularity Model, Demographic-based Filtering Systems, Knowledge-based Recommendation Systems, Content-based Filtering Systems, Collaborative Filtering Systems, and Hybrid Recommendation Systems; after studying these systems,

II. LITERATURE SURVEY

This study was constructed based on the approach to minimizing the learner's efforts and time in searching for the right course. The survey conducted by Viet Anh Nguyen [1] focused on creating a system that will recommend suitable classes for every student in the upcoming semesters based on their current academic scores. They used a variety of data mining and learning analytics techniques to forecast students' learning outcomes for the forthcoming semester and created a model to determine the best courses for each student. They proposed that Each course can be considered an item in a competency matrix, and the students' grades can be considered users are rating the relevant items. We assume that each student's grades are similar, which explains their resemblance. Based on the similarities between students, the User-Based Collaborative Filtering approach forecasts a student's course grade. Their grades in their courses tell us how similar these students are to one another. The degree of resemblance increases with decreasing score discrepancies.

In another study by Sunita B Aher [2], they compare various combinations of data mining algorithms, like clustering and association rule algorithms, association rule mining of classified and clustered data, combining clustering and classification algorithms into association rule algorithms, and solely association rule algorithms. They consider the algorithms for ADTree classification, Simple K-means clustering, and the Apriori association rule. They contrast various combinations of data mining algorithms, such as clustering and association rule algorithms, association rule mining of categorized and clustered data, integrating clustering and classification algorithms into association rule algorithms, and solely association rule algorithms. According to their simulation, the best combination of algorithms is the combination of clustering, classification & association rule mining.

This study by Jing Li[3] looks at how to apply customized recommendation technology, which is extensively used in the business world, to online learning. The platform for customized learning based on a collaborative filtering algorithm is then built and used. Combining the data from the data processing services with the model from the model library, executing algorithm calculations in accordance with the algorithm formula, and then proposing the things customers need are how the personalized recommender system achieves its goal. One must first compute similarities between computed users or objects to locate similar users or items, i.e., neighboring users. Next, it predicts scores by averaging the scores of adjacent users.

Similar to the study done by Viet Anh Nguyen [1], Huynh-Ly Thanh-Nhan [4] also proposed a system with three main feature groups: grading prediction, transferring data, and course recommendation. Training/Predicting application was implemented in terms of desktop application and pre-processing the missing data features/values. After training, the system transfers the grading matrix table from app-server to web-server. After predicting, all grades are stored in the grading matrix and transferred to a web application for course recommendation.

In the paper [5] by Jinjiao Lina Through the addition of expert information and sparseness regularisation in the computation, they proposed a sparse linear-based technique for top-N course recommendation. The method they suggested primarily focuses on the accuracy of course recommendations compared to the empirical data they gathered from experts.

In another web-based system by Ko-Kang Chu [6] Through the course selection method, actual course selection records for two classes across two academic years are gathered. The order of the students' preferences was established by their recommendation process, Then most appropriate courses can subsequently be selected for recommending learners.

In paper [7], this collaborative recommender system used collaborative filtering with the Developing a curriculum (DACUM) Method. This method recommends the course to learners on their knowledge level, learning skills, and learner profile. Data from nine classes were collected for this method. This recommendation system has a 0.6 higher mean value satisfaction than a normal collaborative filtering recommender.

In the study [8] where all types of recommended systems were studied and their challenges were identified. They proposed a system that automatically finds similar students and then applies an association rule mining algorithm to their courses to create course association rules.

Discovered courses association rules are used to get recommendations.

Since our major focus was to make an effortless learning experience for learners and most papers were focused on university students and how to suggest better courses, we studied [9],[10], which were based on surveys and research on the student learning experience. In a paper [9] by Shivangi Dhawan, The SWOC analysis was conducted to understand various strengths, weaknesses, opportunities, and challenges associated with the online learning mode during this critical situation. The research tool used for analyzing the data, which was amassed from different sources for this study, is content analysis, and the research method is descriptive research.

Another study [10] associated with online learning experience during a pandemic The data was collected and analyzed to detect the bottlenecks in online learning and suggestions were given for solving some challenges, which helped to provide the solution not only during a pandemic but also in the long run. Collected the ratings from available e-learning platforms, and a net promoter score (NSP) was generated, which helped in solutions that were enacted with Psychological and biological factors that have been taken into consideration. Also, online learning is the best bet left to counter the current situation but due to the development of such a platform needs to be enhanced to consider methods of learning completely fruitful.

III. PROPOSED SYSTEM AND ARCHITECTURE

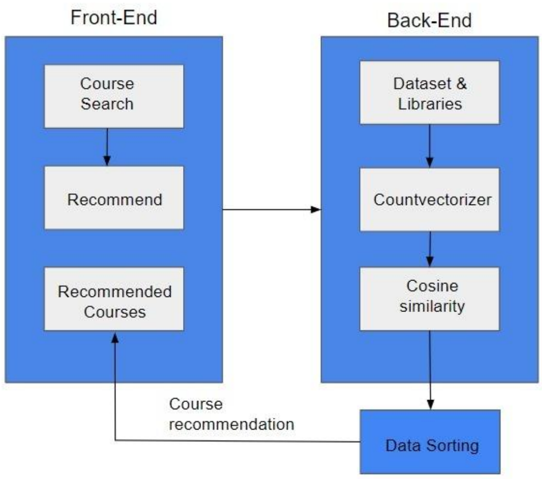

Fig 1 Proposed System and Architecture

FutInitially, we loaded the required libraries such as pandas for deal with data frames and datasets, neattext library removing the stopwords which will be used at data preprocessing step also for fast data processing we had used vectorizers such as countVectorizer and TFIDF vectorizer libraries, and last but not a minor cosineSimilary library for deal data from countVectorizer and logistic regression for training, testing and predicting the course or suggesting the course completing the requirement for the initial phase. As we had to deal with machine learning, we initiated data/text preprocessing, where we cleaned the dataset with as mentioned neattext library. In this process, we first removed the stopwords, punctuations, and then special characters, if any. For building a machine learning course recommendation system with the help of text data, we need to convert it into its vectorized form. For vectorization, we had two options use count vectorization or use (Term Frequency-Inverse Document Frequency)TF-IDF.

ISSN No:-2456-2165

While dealing with countVectorizer, we first created a countVectorizer model and then used fit_transform to fit and transform the model. During the process of fitting, we compute the mean and standard deviation for a given feature which can be used for feature scaling, which is done with the help of the transform method. The output we get is a sparse matrix which contains critical information about the features which will be helpful for future course prediction. Now when a user enters the text about the course he/she wants to learn, this text will be extracted and again preprocessed, cleaned, and vectorized. Now, this count vectorizer data and the course datasets count vectorizer sparse matrices are passed to a function of cosine_similarity_matrix. This method uses the formula

$$\cos \theta = U.V / |U| \times |V|)$$

This function throws an output of their respective values. Which is caught, and the data are sorted in descending order.
Note - These data will range in [-1,1], which is the range of cos θ.

According to the logic, those whose values are close to 1 represent a high similarity, so after we sort the data, we extract the courses with the highest recommendation. At last, we export the data on recommended practices. This was the case if we had used a countVectorizer, but if we used the TFIDF Vectorization method, we would have extracted the dependent and independent variables and then split the data into train and text then with the ratio of 70% training data and 30% testing data with some random state. These training data will be used to train the logistic regression model, which will use the formula (1/ 1 + e^x), which works in such a way that it keeps the data in a range of 0 to 1 where its midpoint is 0.5, so it assumes if the value is greater than 0.5, then the user searched text data is very similar to data set course data and takes it as 1. If the output of logistic regression is less than 0.5, indicated users searched text data is not similar to data set course data and takes it as 0. From the outputs of these classification reports, a confusion matrix is generated, which helps us to visualize the data and get some insights. Finally, the test split data is passed to the model and again checks the testing accuracy where it predicts the course belongs to which subject. In our data, we just had four classifications of subjects - Business Finance, Graphic Design, Musical Instrument, and Web Development. Next, we find the predicted issue in the course data set within the subject column. For all those subjects being matched, we would make a separate table for them and display the result of the predicted course. This model has 96% at the end, and here also, we export the recommended course in a CSV file.

This was the single approach in TFIDF, where one model was used. We have another approach which is called the Pipeline approach. In the pipeline approach, we merge different Transformers (TFIDF Vectorizer) and Estimators (Logistic Regression, MultinomialNB). These models are merged, and a pipeline is created. Then, all the other approaches are similar to single approach modelling. This model has And at the end, here also we export the recommended course in a CSV file.

Fig 2 TFIDF Vectorizer

This is how the course recommendation system model works, and users can choose any methods to implement them and expand the process as per requirement. For example, in the research paper, we implemented the working of countVectorization and integrated it with a web App using streamlit.

IV. FUTURE WORK

For future work, we will be interested in solving the problem of expanding, improving the model, and some other problems such as starting to suggest courses that are more suitable for personal needs, collecting additional information about students (interests, skills, needs) can also be considered.

For more improvements, We can expand our dataset with new popular courses and filters to sort and offer more personalized results for an individual.

V. CONCLUSION

Selection of appropriate courses is one of the important decisions before starting to learn a new skill or upskill for learners. Our project focuses on developing solutions that recommend appropriate and suitable courses for each student based on popular metrics such as reviews and the

ISSN No:-2456-2165

number of subscribers. Our system has proposed models for predicting better courses from the cluster of large data. We have focused on testing two different methods for converting data into vectors for better performance and prediction, which is the basis for making a list of suggested courses for students and learners.

Although the experiments have been conducted with a limited amount of data, the course recommendation based on the most popular courses does show some promise. In this project, we learned about how different factors affect the decision-making of a student while selecting courses, and also about how data can be simplified for better outcomes.

REFERENCES

[1]. Viet Anh Nguyen, et al A course recommendation model for students based on learning outcome" Education and Information Technologies v26 n5, 2021
[2]. Sunita B Aher, et al "Best Combination of Machine Learning Algorithms for Course Recommendation System in E-learning" International Journal of Computer Applications 41(6):1-10, 2012
[3]. Jing Li, et al "Course Recommendations in Online Education Based on Collaborative Filtering Recommendation Algorithm" Complexity (IF2.833), 2020
[4]. Thanh-Nhan Huynh-Ly, et al "Methods for building course recommendation systems" 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE),2016
[5]. Jinjiao Lina, et al "Intelligent Recommendation System for Course Selection in Smart Education" 2017 International Conference on Identification, Information and Knowledge in the Internet of Things,2017
[6]. Ko-Kang Chu, et al "Designing a Course Recommendation System on the Web-based on the Students' Course Selection Records" EdMedia + Innovate Learning, 2003
[7]. Ji-won Han, et al "A collaborative recommender system for learning courses considering the relevance of a learner's learning skills" Cluster Comput 19, 2273–2284, 2016
[8]. Amer Al-Badarenah, et al " An Automated Recommender System for Course Selection" (IJACSA) International Journal of Advanced Computer Science and Applications, 2016
[9]. Shivangi Dhawan, et al "Online Learning: A Panacea in the Time of COVID-19 Crisis" Journal of Educational Technology Systems, 2020
[10]. Thakker, S.V., et al, "Systematic research of e-learning platforms for solving challenges faced by Indian engineering students", Asian Association of Open Universities Journal. 2021