COURSE RECOMMENDATION

BALAJI SHANMUGAVELU S
MSC DATA SCIENCE
THIAGARAJAR COLLEGE OF ENGINEERING
MADURAI-625009
TAMILNADU
INDIA
balajishanmugavelu@student.tce.edu

JOSHIK RAJ S
MSC DATA SCIENCE
THIAGARAJAR COLLEGE OF ENGINEERING
MADURAI-625009
TAMILNADU
INDIA
joshik@student.tce.edu

***Abstract*— Have you ever struggled to search for a course online? Do you want to get the best course among all courses? With the vast number of courses available on platforms such as Udemy, Coursera, and Edx it can be overwhelming for individuals to navigate and select the right course for them. The typical method of choosing courses is a time-consuming, and laborious activity that not only hurts performance but also negatively impacts the learning experience. So, selecting the right courses during formative years could aid in professional growth more effectively. To address these challenges, course recommendation systems are essential to suggest appropriate courses that boost both performance and happiness levels. This study proposes a course recommendation system that utilizes advanced machine learning algorithms to analyze users' choices and generate highly personalized and accurate course recommendations. The proposed recommender system would function better by addressing the shortcomings of the most fundamental individual recommender systems. The system is evaluated using real-world data and the results show that it can generate high-quality recommendations that are well-aligned with student needs and preferences. Overall, the proposed course recommendation system has the potential to significantly improve the learning experience for students and help them to achieve their academic goals.**

Keywords—Content-based Filtering, K-means, Cosine Similarity, Linear Similarity.

I. INTRODUCTION

Course recommendations are suggestions for courses that are tailored to individual needs and interests. These systems primarily aim to improve the teaching and learning process while maintaining a high standard of education Researchers today are working to make personalized course searching even more effective. The user may be given decision-making abilities by these techniques. A recommender system's graphical user interface and design are tailored to the applications it will be used for based on the products it will suggest. The development of recommender systems was sparked by the notion that people frequently consult others when making important decisions in life. As Students may not be aware of all the courses available to them, which can make it difficult for them to make an informed decision and some students may have false assumptions about certain courses, which can affect their decision-making. Students may feel pressured to choose a course that is popular or in high demand, even if it may not align with their interests or strengths. As there are several popular MOOC platforms including Coursera, edX, Udemy, and Khan Academy. The difficulty in those platforms is there are more courses to choose from and it can be difficult to know where to start and find the right course. Using content-based filtering techniques, it was the goal of several projects to recommend similar courses and also the top courses based on the ranking to improve their knowledge. The complexity of recommendations has increased over the past few years. It is crucial to suggest useful courses in addition to choosing passable ones. The relevance of courses was assessed based on their importance across all university disciplines, their degree of interconnectedness, and their relevance to the student's field of study. From that moment, a set of courses will be recommended based on the user's choice from Udemy courses.

II. LITERATURE SURVEY

NEW SYSTEM	EXISTING ONE
Collaborative- based filtering is not used.	Collaborative Filtering is used.
K-Means is used.	K-Means is not used.
Recommendation is based on the user's choice.	Recommendation is based on the past history of the end user.

Figure 1: COMPARISON BETWEEN NEW AND EXISTING ONE

.(O'Mahony & Smyth,(2007).

- The drawback of the state of the Art in Methodologies of Course Recommender Systems is, to improve user interaction, accurately understand user preferences and provide more personalized recommendations. This problem is rectified in this recommended system. (Guruge, D. B., Kadel, R., & Halder, S. J. (2021))

- A Course Recommender System based on Graduating Attributes and the proposed algorithm needs to be tested on real data to further validate its effectiveness. (Bakhshinategh, B., Spanakis, G., Zaiane, O., & ElAtia, S. (2017))

- Intelligent recommendation system for course selection in smart education. The drawback is no proper testing is given. (Lin, J., Pu, H., Li, Y., & Lian, J. (2018))

- A course recommender system for University College Dublin's online enrolment application uses a collaborative filtering algorithm. The drawback of this paper is that didn't develop an algorithm with more features but in this course recommendation system, there will be more features trained to get accurate recommendations. (O'Mahony & Smyth,(2007).

- In an Automated Recommender System there is no comparison between their method and other typical methods. (Al-Badarenah, A., & Alsakran, J. (2016)).

- Consideration of recommendation systems that begin with recommending the qualification path. (Maphosa, M., Doorsamy, W., & Paul, B. (2020)).

- Testing the result on a huge amount of data. This drawback is rectified by using the udemy dataset. (Aher, S. B., & Lobo, L. M. R. J. (2012)).

- Used Atomization of the combination of ADTree classification and Apriori Association Rule algorithm. (Aher, S. B., & Lobo, L. M. R. J. (2012)).

- Personalized course recommendation graph: Course dependency graph (using Mann-Whitney U-test) and Grade prediction using Collaborative (Backenköhler, M., Scherzinger, F., Singla, A., & Wolf, V. (2018)).

- Cluster students on demographic data and previous course scores and association rules are mined from the clusters to do the recommendations (Asadi, S., Jafari, S., & Shokrollahi, Z. (2019)).

- CFRS with K-means to cluster students, KNN to select K-most similar groups, ARM to discover course selection patterns (Al-Badarenah, A., & Alsakran, J. (2016).

III. DATASET

This study is focused on evaluating the effectiveness of a course recommendation system using Udemy courses as the sample. The Udemy course dataset used for training as well as testing was obtained from Kaggle (Give URL). It contains information on the number of courses available on Udemy and also the course details. This dataset includes 999+ records from Udemy courses in the following 4 categories: business finance, graphic design, musical instruments, and web design.

Feature	Description
Dataset Name	Udemy Course Dataset
Size	1000 records
Number of Features	7
Data Type	categorical, numerical
Target Variable	Course Title
Data Source	Udemy
Date of Creation	2021

Figure 2: DATASET DESCRIPTION

IV. TECHNIQUES USED FOR PRE-PROCESSING:

❖ **Feature Extraction**

The process of turning raw data into numerical features that can be processed while keeping the courses in the original data set is known as feature extraction. Compared to using machine learning on the raw data directly, it produces better results.

❖ **CountVectorizer**

Here, CountVectorizer is used because it is a great tool provided by the sci-kit-learn library in Python. It is used to transform a given text into a vector based on the frequency (count) of each word that occurs in the entire text.

❖ **TfidfVectorizer**

The TfidfVectorizer creates a matrix of TF-IDF characteristics from a group of unprocessed documents. Fast Text and Word2Vec Word Embeddings Python Implementation. The simplicity and ease of usage of TF-IDF are its major benefits. It is simple to calculate, it is computationally cheap, and it is a simple starting point for similarity calculations (via TF-IDF vectorization + cosine similarity).

- **Removing Duplicates and Special Characters**

  This is used for removing unwanted characters like whitespace and unwanted punctuation. By using this the course will be recommended with a suitable title.

V. ALGORITHMS USED

K-Means:
This algorithm is used to group the unlabelled dataset into different clusters. Therefore It allows us to cluster the data into different groups and is a convenient way to discover the categories of groups in the unlabelled dataset on its own without the need for any training. It is an iterative algorithm that separates the unlabeled dataset into k distinct clusters, each of which contains only one dataset and shares a set of characteristics.

Purpose of K-means clustering for this course recommendation:
The purpose of using K-means clustering in course recommendation systems is to group similar courses based on certain features, such as course content, instructor, student feedback, and more. By clustering courses into similar groups, a recommendation system can more easily match a user's preferences and interests to courses that are likely to be of interest to them. For example, if a user is interested in courses on data science, the recommendation system can identify the cluster of courses on data science and recommend the top courses within that cluster to the user.

Steps for recommendation:
Step 1: Pick K to get the number of clusters.

Step 2: Pick K points or centroids at random. (It might not be the input dataset.)

Step 3: Assign each data point to its nearest centroid, which will create the K clusters that have been predetermined.
Step 4: Determine the variance and relocate each cluster's centroid.

Step 5: Re-assign each data point to the new centroid of each cluster by repeating the third step.

Step 6: Go to step 4, if there is a reassignment; otherwise, go to FINISH.

Step 7: The finished model.

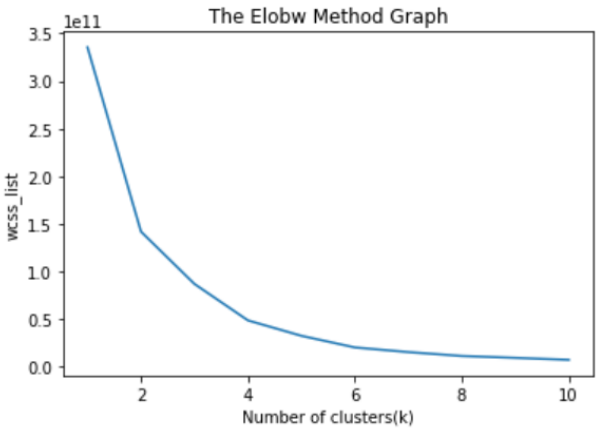

Figure 3: K-Means Representation

Through k-means there is a graph plot that will show the price and subscription of the courses which is given below:

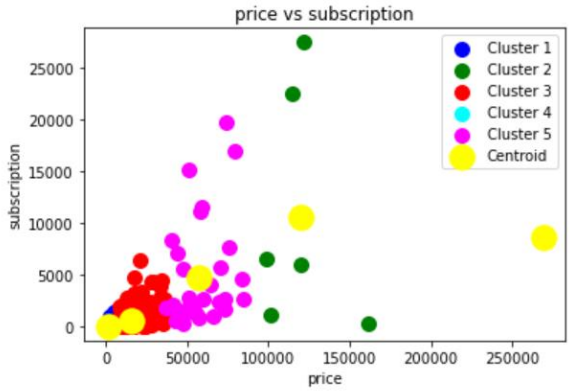

Figure 4: price vs subscription

The price vs subscription plot can be used in this course recommendation system to visualize the relationship between the cost of a course and the number of students who have subscribed to it. This information can be used to make informed recommendations to students by considering both their interests and their budget constraints. For example, if the plot shows that students are more likely to subscribe to less expensive courses, the recommendation system can prioritize recommending lower-cost courses to students who have budget constraints. On the other hand, if the plot shows that the cost of a course does not significantly affect the number of students who subscribe to it, the recommendation system can focus on recommending courses based solely on the student's interests and preferences. In summary, the price vs subscription plot can provide valuable insights into the purchasing behavior of students, which can be used to make more informed course recommendations.

Content-based filtering algorithm:

One widely used method of recommendation or recommender systems is content-based filtering. This recommendation system makes recommendations for you based on the course title and its keywords. Here we used content-based filtering to group courses according to certain

keywords, discover the customer's preferences, look up those phrases in the dataset, and then suggest related courses.

Logistic Regression:

- Logistic regression is used in this course recommendation system to predict the likelihood that a user will take a particular course.
- Logistic regression models the relationship between the independent variables (user characteristics and course features) and the target variable.
- The logistic regression model outputs a probability between 0 and 1, which can be used to predict the likelihood that a user will take a particular course.
- Logistic regression can be used to make personalized course recommendations to users based on their characteristics and preferences.
- The objective is to increase the likelihood of user engagement with the recommended courses.

Naïve Bayes:

It primarily uses a huge training set for text classification. As it is a probabilistic classifier, which means it predicts based on the probability of an object.

$$\left(\frac{A}{B}\right) = P\left(\frac{B}{A}\right) P(A)/P(B)$$

Where,

P(A/B) is the probability that the student is interested in the course given their features (e.g. past course preferences, grades, demographics, etc.)

P(B/A) is the probability of observing the student's features given that they are interested in the course

P(A) is the prior probability of the student being interested in the course, estimated based on the behavior of all students in the system

P(B) is the probability of observing the student's features, which is a normalization constant.

In the context of a course recommendation system, Naive Bayes can be used to predict the likelihood of a student being interested in a particular course based on their past behavior and preferences.

Confusion Matrix for naïve bayes:

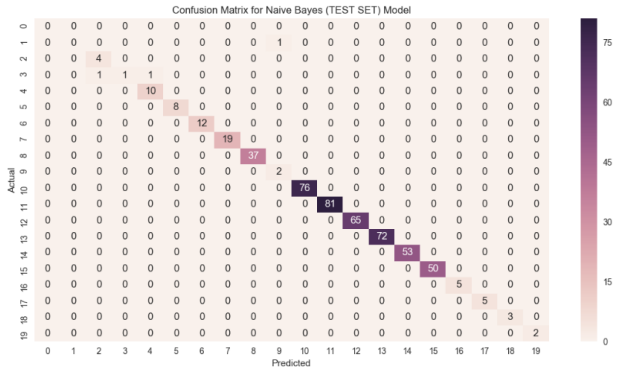

Figure 5:confusion matrix for naive Bayes

Accuracy, precision, and recall:

Logistic Regression	Naïve Bayes	K-Nearest Neighbours
37.008%	99.409%	99.213%
37.008%	99.409%	99.213%
0.147	0.912	0.906

Figure 6:Accuracy, Precision, Recall

- The accuracy, precision, and recall are calculated to represent which algorithm shows the better results for a recommendation.
- From the above results we can infer that the best algorithm to cluster the course to recommend is k-means.

VI. SIGNIFICANT RESULTS

This course recommend system provides an accuracy of 99% whereas the other model gives only 75.11% accuracy. (Gulzar, Leema, Deepak-2018). Also, it says about the interest of the student in a particular course based on their past behavior and preferences. The recommendations can be the user's choice i.e., the user can able to give several recommendations based on the similarity scores. From this, we can infer that this model gives the best recommendation to the end user.

VII. OUTPUT

In the output, this paper will give you better recommendations as user preferences, and also the best thing is, it will recommend the number of recommendations the users need for that particular course. So, based on the course title and also a similarity of the course will be displayed to the user.

```
Enter the Course you are interested:How To Maximize Your Profits Trading Options
Enter the number of recommendations you need:5
```

	course_title	similarity-scores
408	Trading Options Basics	0.577350
43	Options Trading - How to Win with Weekly Options	0.566947
94	Intermediate Options trading concepts for Stoc...	0.530330
136	Forex Trading with Fixed 'Risk through Options...	0.530330
193	Trading Options For Consistent Returns: Option...	0.530330

Figure 7: Output

VIII. CONCLUSION

This recommender system is employed to remind students to complete their assignments, caution them against challenging courses, and recommend courses that could be of benefit to them. As a result, the system supports students' decision-making during the enrollment process at the start of every semester.

We created three algorithms exclusively for the course suggestion. The first algorithm looks for the courses with the highest enrollment rates. The second method makes use of student commonalities based on the courses they are interested in. The third algorithm makes course suggestions.

The first disclosed algorithm was the one that was best appropriate for the selected course suggestion. Most often, students choose the simpler courses that were included in their templates. On the other hand, the second algorithm that took use of students' commonalities produced the greatest results for the suggestion of optional courses. However, given how highly students rated the recommendations, we chose to use every approach in the system. Additionally, we only advised elective courses if we thought students would be able to complete them and if there were open spaces in their schedules.

Accordingly, this paper will suggest a set of courses based on how many suggestions the user needs. The user may choose any of those courses from that list and decide which one would help them move in a better route depending on their preferences.

IX. LIMITATIONS

The further work of this paper is to design a website for society. So, if this turned into a website it will be more user-friendly as it can be designed to show the course details like if the course is paid or unpaid, and also the ratings to the user who needs the course recommendation for their purposes. Recommendation systems can become computationally expensive as the number of users and items increase. As a result, it may become necessary to use more advanced techniques, such as distributed computing or matrix factorization, to ensure the system remains scalable.

X. REFERENCES

[1] Gulzar, Z., Leema, A. A., & Deepak, G. (2018). Pcrs: Personalized course recommender system based on a hybrid approach. *Procedia Computer Science*, *125*, 518-524.

[2] O'Mahony, M. P., & Smyth, B. (2007, October). A recommender system for online course enrolment: an initial study. In Proceedings of the 2007 ACM conference on Recommender systems (pp. 133-136).

[3] Al-Badarenah, A., & Alsakran, J. (2016). An automated recommender system for course selection. International Journal of Advanced Computer Science and Applications, 7(3), 166-175.

[4] Parameswaran, A., Venetis, P., & Garcia-Molina, H. (2011). Recommendation systems with complex constraints: A course recommendation perspective. ACM Transactions on Information Systems (TOIS), 29(4), 1-33.

[5] Guruge, D. B., Kadel, R., & Halder, S. J. (2021). The state of the art in methodologies, of course, recommender systems—a review of recent research. Data, 6(2), 18.

[6] Thanh-Nhan, H. L., Nguyen, H. H., & Thai-Nghe, N. (2016, October). Methods for building course recommendation systems. In 2016 Eighth International Conference on Knowledge and Systems Engineering (KSE) (pp. 163-168). IEEE.

[7] Sulaiman, M. S., Tamizi, A. A., Shamsudin, M. R., & Azmi, A. (2020). Course recommendation system using the fuzzy logic approach. Indonesian Journal of Electrical Engineering and Computer Science, 17(1), 365-371.

[8] Bakhshinategh, B., Spanakis, G., Zaiane, O., & ElAtia, S. (2017, April). A course recommender system based on graduate attributes. In International Conference on Computer Supported Education (Vol. 2, pp. 347-354). SCITEPRESS.

[9] Lin, J., Pu, H., Li, Y., & Lian, J. (2018). Intelligent recommendation system for course selection in smart education. Procedia Computer Science, 129, 449-453.

[10] Maphosa, M., Doorsamy, W., & Paul, B. (2020). A review of recommender systems for choosing elective courses. International Journal of Advanced Computer Science and Applications, 11(9), 287-295.

[11] Aher, S. B., & Lobo, L. M. R. J. (2012). Combination of clustering, classification & association rule-based approach for course recommender system in E-learning. International Journal of Computer Applications, 39(7), 8-15.

[12] Aher, S. B., & Lobo, L. M. R. J. (2012). Mining association rule in classified data for course recommender system in e-learning. International Journal of Computer Applications, 39(7), 1-7.

[13] Backenköhler, M., Scherzinger, F., Singla, A., & Wolf, V. (2018). Data-Driven Approach towards a Personalized Curriculum. International Educational Data Mining Society.

[14] Asadi, S., Jafari, S., & Shokrollahi, Z. (2019). Developing a course recommender by combining clustering and fuzzy association rules. Journal of AI and Data mining, 7(2), 249-262.

[15] Al-Badarenah, A., & Alsakran, J. (2016). An automated recommender system for course selection. International Journal of Advanced Computer Science and Applications, 7(3), 166-175.