Course Recommendation System Using Machine Learning

[1]Ch.Suneetha, [2]M.Jahnavi, [3]M.Yashwanth, [4]Naonit Kumar Goutam

[1]Associate professor, [2,3,4]Undergraduate student
Department of Electronics and Communication Engineering,
Vignana Bharathi Institute of Technology
Ghatkesar, India.

***Abstract*- A computer-based algorithm known as a Course Recommendation System helps students choose courses based on their unique interests, academic skills, and desired careers. Making the appropriate decision during formative years is crucial since the outcome will affect the future. This method uses machine learning algorithms to create a list of courses that most closely fit the user's tastes by analyzing user activity data, including ratings and past search history. To recommend the best courses based on their interests, machine learning algorithms including stemming, count vectorization, and cosine similarity are used. The main objective of this paper is to lighten the workload of the students while maintaining their attention.**

***Index Terms*- Course recommendation, Machine learning, Stemming, Count vectorization, Cosine similarity.**

I. INTRODUCTION

The field of education is one that is just starting to take off since there are so many new opportunities to learn and grow as the globe develops so quickly. Everyone, especially students, must arm themselves with the most recent technologies because everything is evolving at such a rapid rate. An AI-powered system known as a course recommendation system offers learners individualized course options based on their interests, prior educational experiences, and aspirations. To make recommendations that are specifically tailored to a learner's requirements and interests, the system can assess a variety of data, including academic history, preferred learning style, prior course selections, and other personal traits. This system will save the time of an individual by choosing the insignificant course in which they are not interested instead of selecting the course of their choice. In this system, we are using machine learning along with its techniques such as Stemming, Cosine similarity, and Count vectorization. There are many different types of recommendation systems, of course, including content-based, collaborative filtering, and hybrid systems. A new technology called a course recommendation system can assist individuals and different institutions in selecting and offering the courses they want. A course recommendation system is a type of information filtering system that suggests relevant courses to users based on their interests and preferences. This system uses various techniques such as natural language processing, machine learning, and data mining to analyze the user's behavior and provide personalized recommendations. Cosine similarity, stemming, and count vectorization are commonly used techniques in course recommendation systems.

Cosine similarity is a measure of similarity between two vectors in a high-dimensional space, which can be used to compare the similarity between course descriptions. Stemming is a technique used to reduce words to their root form to improve the efficiency of the system. Count vectorization is a method used to convert text data into numerical vectors, which can be used for further analysis.

II. LITERATURE SURVEY

[1] Behdad Bankshinategh, Gerasimos Spanakis, Osmar Zaiane, and Samira ElAtia (Jan 2017) proposed a course recommendation system for students based on the assessment of their “graduate attributes”. They described how the assessment of the GAs can be used to generate course recommendations. The algorithm relies on assessment data provided by the students and works in a collaborative filtering context taking into account the time factor.

[2] M. Rekha Sundari, Geetha Shreya, and T. Jawahar (Nov 2020) proposed a recommendation system to deal with random choices that do not take into consideration the chosen course level. This recommender system based on type classification rules (IF-THEN), adopts a methodological approach that is based on the student's performance in previous semesters and the student's choice of electives.

[3] O'Mahony, Michael P., and Barry Smyth (Oct 2007) developed a course recommender system for University College Dublin's online enrollment application. They outlined the factors that influence student choices and propose solutions to address some of the key considerations that are identified. Also, empirically evaluated their approach using historical student enrolment data and show that promising performance is achieved with their initial design.

[4] A Hybrid Course Recommendation System by Integrating Collaborative Filtering and Artificial Immune Systems by Pei-Chann Chang, Cheng-Hui Lin, and Meng-Hui Chen (July 2016). This research proposed a two-stage user-based collaborative filtering process using an artificial immune system for the prediction of student grades, along with a filter for professor ratings in the course recommendation for college students. They tested for cosine similarity and Karl Pearson (KP) correlation in affinity calculations for clustering and prediction. higher mean grades.

III. PROPOSED SYSTEM

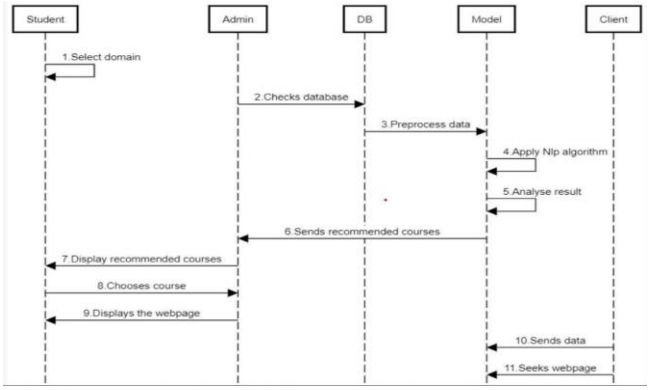

Fig.1 Sequence Diagram

The proposed approach eliminates the suggestion of deceptive courses. The Cosine Similarity and Stemming Method are helpful in recommending a path. By counting the most common features shared by the products, the Cosine Similarity is frequently used to match similar product item sets. It calculates the cosine of the angle formed by two vectors that are projected onto a multidimensional space. A stemming algorithm is used during the process to reduce the tokens to the root token.

Cosine Similarity Algorithm

Regardless of size, cosine similarity is a statistic that aids in identifying how similar data objects are. You can use cosine similarity to measure the similarity between two sentences in Python. The dataset's data objects are treated as vectors by cosine similarity. It is useful because the angle between two identical data objects can still be modest even though their sizes are separated by the Euclidean distance. The degree of resemblance increases with decreasing angle. The angle of the data object, not its size, is captured when it is represented in multidimensional space.

$$Cos(x, y) = x \cdot y / ||x|| * ||y|| \quad (1)$$

where,

- x . y = product (dot) of the vectors ‘x’ and ‘y’.
- ||x|| and ||y|| = length of the two vectors ‘x’ and ‘y’.
- ||x|| * ||y|| = cross product of the two vectors ‘x’ and ‘y’.

Stemming Algorithm

The process of stemming involves breaking a word down into its basic parts. Suffixes, prefixes, and other headwords are stems that are attached to other stems. Natural Language Processing (NLP) and Natural Language Understanding (NLU) both benefit from stemming. AI and stemming expertise are used to glean useful information from enormous sources like big data and the Internet. You will find more results when you search for and find additional word formats. In many instances, using the headword—the word's fundamental morphological form—will yield the greatest results. Stemming is done via an algorithm that either people or AI systems can employ to find the headword. Stemming employs a variety of techniques to strip words of their varying inflections.

Count Vectorization

Count Vectorization is the process of breaking a sentence or text into words by counting, changing all words to lowercase, and deleting special characters. This text data needs to be vectorized, punctuated, and converted to lowercase because the NLP model only understands and accepts numerical data. A recognized word vocabulary is created and will be used later. It encrypts secret text and returns the size of the full vocabulary as well as an integer indicating how many times each word appears in the document's encoded vectors.

IV. DATASET INTERPRETATION

This paper uses a dataset (Coursera Dataset) taken from Kaggle. The dataset consists of various courses, their difficulty level, rating, university, etc. It contains over 3522 rows with 7 columns. The column names are: Course Name represents the name of the course, University represents the name of the university providing the particular course, and Difficulty Level represents the difficulty level ranging from beginner, intermediate, and advanced. Course Rating represents opinions ranging from 1-5 whereas Course Description gives an overview of the course. Skill represents the skills required for the course. These dataset parameters help a student finalize the course based on his/her interest.

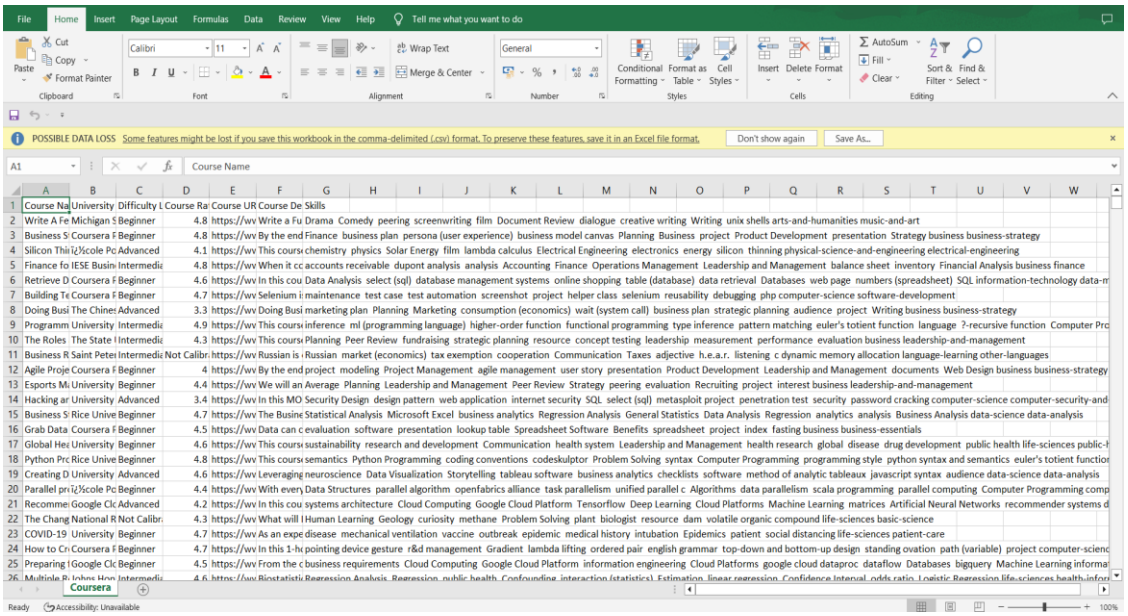

Fig.2 Dataset Used

V. EXECUTION

As the initial step, we dump the required dataset on Google Colab platform. Also, display the attributes of the dataset. The next step is Data Visualization in which we represent difficulty level of the courses and ratings of them through a pie chart and a bar graph respectively. Afterwards, we try to replace brackets, colons, and spaces with commas in the column 'Data Description' after removing spaces between the words in 'Data Course' and 'Data Description' columns. Next, we create a new data frame and rename the attributes by adding an underscore. Count vectorization is our following step in which unnecessary English words will not be counted. Stemming process being the succeeding stage reduces word inflections because we know that Inflection is the process by which words are changed to convey many grammatical categories such as tense, case, voice, aspect, person, number, gender, and mood. Therefore, a word can exist in multiple inflections, but multiple inflections in the same text add redundancy to the NLP process. Therefore, stemming is used to reduce the course names to their basic form or stem.

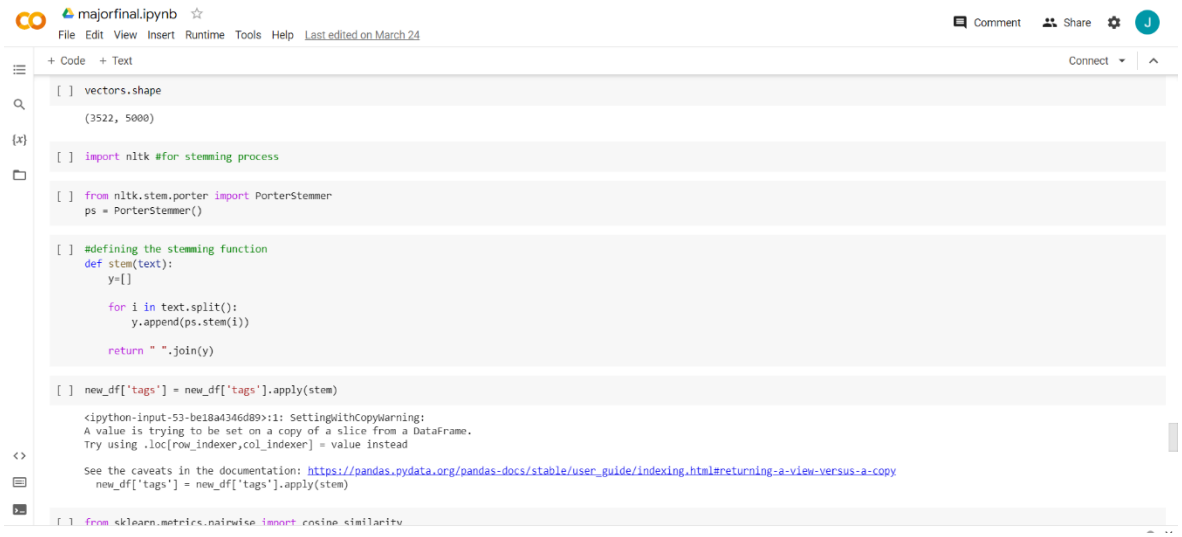

Fig.3 Stemming process

In the next step, Cosine Similarity is applied to the vectors obtained from the stemming process.

Fig.4 Cosine similarity

Once the overall procedure is done, we are able to display recommended courses of desired domain. After this, we import Pickle. For user interface, we create a webpage on Visual Studio Code. In the web code, we import .pkl (pickle) files and define recommend function which takes selected courses argument and return recommended courses. Finally, we create a search box for the selection of courses from the list and also, a button to show the recommended courses. When it is clicked by a user, the recommend function is called.

VI. RESULT

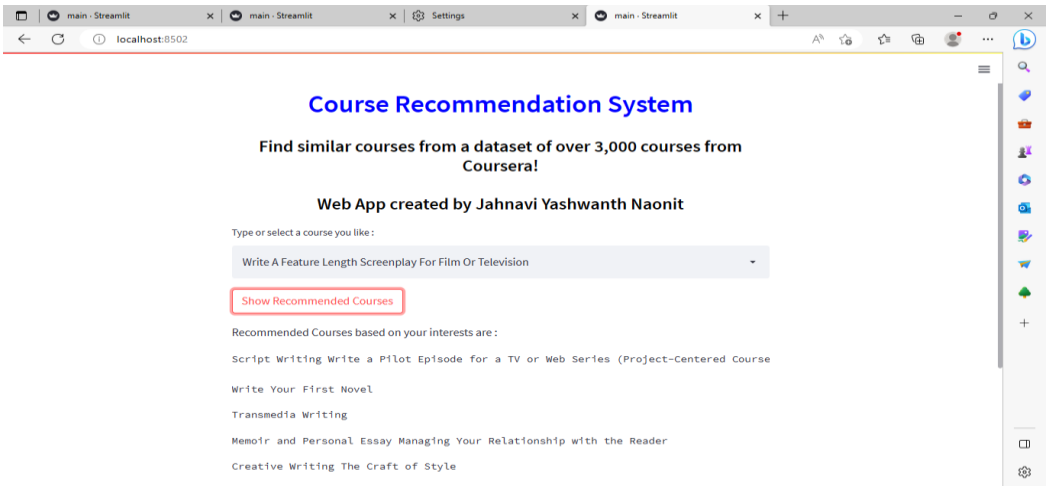

Fig.5 Web page 1

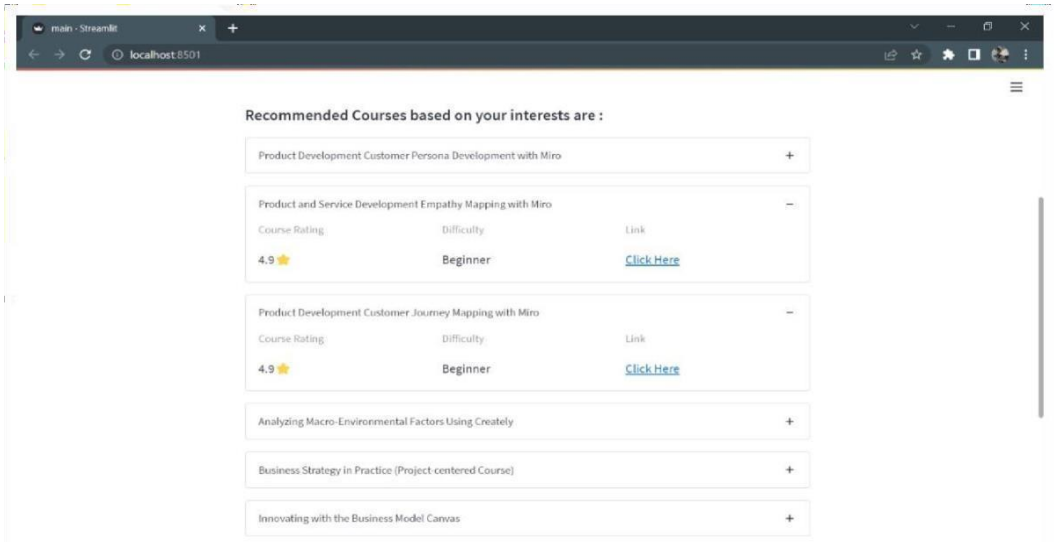

Fig.6 Web page 2

The web page shows the recommended courses according to the domain entered by user. When we click on the "show recommend courses" button, the recommended courses get displayed. On clicking any course from the displayed course list, the difficulty level, rating and link to the respective course gets displayed. By clicking on the link, the page is redirected to the specific course page which the user requires.

VII. CONCLUSION

By using this paper application, we can overcome the problem we are facing in the current situation about the dump of courses. This application's aim is to quickly make sure to find the correct course for the student. It is most efficient in an emergency situation. Finding the right course at the right time can lead to less amount of time wasted and more interest in the student about the course domain.

REFERENCES:

1. Behdad Bakhshinategh, Gerasimos (Jerry) Spanakis, Osmar R. Zaïane and Samira Elatia, "A Course Recommender System based on Graduating Attributes". 9th International Conference on Computer Supported Education, 2017.
2. M. Rekha Sundari, Geetha Shreya, and T. Jawahar, "Course Recommendation System". International Journal of Computer Applications, 175(29): p.13-16, November 2020.
3. MICHAEL P. O'MAHONY AND BARRY SMYTH, "A RECOMMENDER SYSTEM FOR ON-LINE COURSE ENROLMENT: AN INITIAL STUDY". ASSOCIATION FOR COMPUTING MACHINERY, P.133–136, OCTOBER 2007.
4. Pei-Chann Chang, Cheng-Hui Lin, and Meng-Hui Chen, "A Hybrid Course Recommendation System by Integrating Collaborative Filtering and Artificial Immune Systems". Metaheuristic Algorithms in Optimization and Applications, 22 July 2016.
5. Kiratijuta Bhumichitr, Songsak Channarukul "Recommender System for University elective course recommendation" 2017 14th International Joint Conference on Computer Science and Software Engineering (JCSSE)
6. Ahmad Al-Aiad, Rehab Duwairi "A collaborative recommendation system for online courses recommendations" 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML)
7. Viddhesh Sankhe, Janice Shah,Tejas Paranjape, Skilled Based Course Recommendation 2020 IEEE International Conference on Computing, Power and Communication Technologies (GUCON)