Adaptive Learning Paths in Personalized Education: Utilizing Recommender Systems and Cosine Similarity

Gousia Hazra Anjum khan[1], Aashutosh Nishad[2], Kalyani Netam[3] , Jasmine Kispotta[4]

Assistant Professor, Dept. of Computer Science and Engineering, GEC Raipur, CSVTU Bhilai
Student, Dept. of Computer Science and Engineering, GEC Raipur, CSVTU Bhilai
Student, Dept. of Computer Science and Engineering, GEC Raipur, CSVTU Bhilai
Student, Dept. of Computer Science and Engineering, GEC Raipur, CSVTU Bhilai

---***---

Abstract - *Traditional learning methods have been used for years but often struggle to meet the diverse needs and preferences of individual students, decreasing learning efficiency and affecting overall outcomes. Employing the same approach for all students is not advisable. In today's digital age, students have access to abundant learning resources, yet each student's unique requirements, influenced by their mental state and background. Here personalization comes into play. Personalized learning systems that have emerged as a solution to address the diverse needs and preferences of students in a classroom. This research paper presents an approach to personalized learning using Natural Language Processing (NLP) and Machine Learning (ML) algorithms. The proposed system takes user details to generate tags and utilizes collaborative and content-based filtering techniques to recommend study content according to student preferences. By this we can improve the students' overall performance and give them a better career.*

Key Words: Personalized Learning System, Natural Language Processing, Cosine Similarity, Adaptive Learning, Tag based Recommender System

1.INTRODUCTION

In the digital education era, personalization and adaptive approaches are emerging to focus on individual learners' unique needs and preferences. Empowered by advancements in Natural Language Processing (NLP) and Machine Learning (ML) algorithms, personalized learning systems promise to revolutionize traditional education by providing tailored experiences that adapt to each learner's pace, interests, and styles. Leveraging learner and resource data, these systems deliver targeted recommendations, adaptive content delivery, and real-time feedback to enhance engagement, motivation, and learning outcomes.

1.1 Significance of Personalized Learning

Traditional one-size-fits-all educational models often struggle to accommodate the diverse needs, backgrounds, and learning trajectories of individual learners. Personalized learning, on the other hand, 1 recognizes and embraces this diversity, offering flexible, learner-centered approaches that empower learners to take ownership of their learning journey. By providing customized pathways, adaptive resources, and targeted interventions, personalized learning systems have the potential to unlock each learner's full potential, cultivate lifelong learning habits, and bridge achievement gaps across diverse learner populations. From research we learned that Personalized learning (PL) has emerged as an effective practice to address the diverse needs of learners in recent years. Many definitions have been published to define PL by government offices, educational policy organizations, educational foundations and initiatives, influencers, and researchers. The most commonly referenced definition was provided in 2010 by the US Department of Education Office of Educational Technology, which defined personalization as "instruction that is paced to learning needs, tailored to learning preferences, and tailored to the specific interests of different learners" (US DOE, 2010). The majority of personalization efforts were centered around identifying and accommodating students' "interests" and "needs," though few additional details were offered to operationally define these terms. Definitions included myriad design approaches to accommodate learner characteristics, including pace, delivery approach, coverage, and sequence of instruction, as well as methods of scaffolding, delivering, and assessing mastery of content. The learner outcomes that personalized learning could target spanned motivation, skill, and achievement, and not all definitions clearly defined an aim. Perhaps the most salient feature of this thematic representation of personalized learning was the complexity endemic in the definitions. With the exception of a very general definition provided by Cuban (2018), every definition included more than one learner characteristic, design component, and/or learner outcome. This suggests that implementations of personalized learning are likely to be complex, where the effects of multiple design factors may need to be parsed or interacted, and parallel analyses may need to be conducted to examine effects on discrete variables among those targeted in a design. This complexity induces challenges for the systematic study of personalized learning, as enacted in authentic educational settings. Figure from Bernacki, Matthew & Greene, Meghan &

Lobczowski, Nikki. (2021). A Systematic Review of Research on Personalized Learning: Personalized by Whom, to What, How, and for What Purpose(s)?. Educational Psychology Review. In the above figure, the yellow line shows that by allowing students to choose contents on their own, their motivation increases and the green line shows that if we or our system provides content to users based on their interests and needs, this will increase their motivation as well as their performance.

1.2 Role of NLP and ML in Personalized Learning

At the heart of personalized learning systems lie NLP and ML algorithms, which enable the intelligent analysis, interpretation, and adaptation of educational content and learner data. NLP techniques empower these systems to extract insights from unstructured text, such as user feedback, learning materials, and communication channels, facilitating the generation of user profiles, topic modeling, and sentiment analysis. ML algorithms, including collaborative filtering, content-based filtering, and reinforcement 2 learning, enable personalized learning systems to dynamically adapt recommendations, predict learner behavior, and optimize learning pathways based on individual learner characteristics and preferences.

1.3 Objectives of the Research Paper

In this research paper, we aim to explore the design, implementation, and evaluation of a personalized learning system that harnesses the capabilities of NLP and ML algorithms to deliver adaptive and tailored learning experiences. Specifically, we will –

● Investigate the integration of NLP techniques for user modeling, content analysis, and sentiment analysis in personalized learning systems.

● Explore ML algorithms for recommendation, prediction, and adaptive learning in the context of personalized education.

● Evaluate the effectiveness of the proposed personalized learning system in improving learner engagement, satisfaction, and learning outcomes through quantitative metrics and qualitative feedback.

2. LITERATURE REVIEW

2.1 Personalized Learning Systems

Personalized learning customizes education to fit individual learners' needs. In "A Systematic Review of Research on Personalized Learning," Bernacki et al. (2021) examine four key aspects of personalized learning:

1. Personalized by Whom: Identifies who customizes the learning experience, whether educators, instructional designers, or AI systems.

2. Personalized to What: Looks at which elements are personalized, like content, pacing, assessments, or learning pathways.

3. Personalized How: Describes the methods used for personalization, such as adaptive algorithms, technology, or manual customization.

4. Personalized for What Purpose(s): Explores the goals, including improved learning outcomes, increased engagement, or learner satisfaction.

The authors conduct a systematic review to explore these questions, highlighting gaps in the research and suggesting directions for future study.

2.2 Recommender Systems

Recommender systems play a vital role in personalized experiences across various industries. In "A Systematic Review and Research Perspective on Recommender Systems," Roy and Dutta (2019) discuss key aspects of these systems, including their algorithms, evaluation methods, and applications. They analyze how these systems are used in different fields, from retail to entertainment, and how they are evaluated for effectiveness. The paper also highlights emerging trends and challenges, emphasizing the need for transparent and unbiased recommendations. The authors suggest research directions to improve recommender systems, making the paper a useful resource for those interested in personalized content delivery.

2.3 Artificial Intelligence in Personalized Education

Artificial Intelligence (AI) is increasingly used to create personalized learning experiences. Fernandes et al. (2022), in "Advancing Personalized and Adaptive Learning Experience in Education with Artificial Intelligence," explore how AI can tailor educational content, instruction, and assessments to individual learners. The paper discusses current AI-driven educational technologies and explores the benefits and challenges of using AI in education. The authors emphasize AI's potential to enhance learner engagement and outcomes while acknowledging ethical concerns and technical limitations.

2.4 Cosine Similarity

Cosine similarity is a measure used to assess the similarity between two vectors, often applied in text analysis and information retrieval. In "Cosine Similarity to Determine Similarity Measure: A Study Case in Online Essay

Assessment," Lahitani et al. (2020) examine its use in online essay grading. The study assesses the accuracy and reliability of cosine similarity for evaluating student essays. The authors describe how cosine similarity scores are calculated and compare them with human-assigned grades to gauge the method's accuracy. The paper explores the potential of cosine similarity in automated essay grading and identifies its benefits and limitations.

3. METHODOLOGY

3.1 Data Collection

Data collection in our system involves two primary components: user details and learning content.

3.1.1 User Details

User details are gathered through registration and profiling, capturing demographic information, learning preferences, and interests. This includes past interactions, such as viewed content, completed courses, ratings, and feedback.

3.1.2 Learning Content

Learning content refers to articles, videos, courses, textbooks, and other materials relevant to user learning goals. We collect metadata for each content piece, including title, description, keywords, rating, and tags. This metadata forms the basis for content recommendation and adaptive learning paths.

3.2 Tag Generation

In a personalized learning system, maintaining tags that accurately represent content is essential for understanding user profiles and providing tailored recommendations. One technique crucial for this task is tokenization. Tokenization involves breaking down text into individual words or tokens, forming the foundational step in preprocessing textual data for further analysis.

"Explore machine learning basics with this beginner-friendly tutorial. Topics include supervised learning, unsupervised learning, regression, classification, and clustering algorithms."

After tokenization, the extracted tags include:

"machine learning" , "beginner-friendly tutorial" , "supervised learning" , "unsupervised learning" , "regression" , "classification" , "clustering algorithms"

Following tokenization, the subsequent steps may include:

Normalization: Convert tokens to lowercase and remove punctuation for consistency.

Stop word Removal: Eliminate common stop words like "with," "this," "and," which do not convey significant meaning.

Stemming or Lemmatization: Reduce words to their root forms to consolidate variations (e.g., "learning" to "learn").

Entity Recognition: Identify domain-specific entities like "machine learning" or "clustering algorithms" for contextual understanding.

Feature Engineering: Create additional features, such as n-grams or thematic categories, to capture nuances in the content.

Vectorization: Convert processed tokens into numerical representations using techniques like TF-IDF or word embeddings for machine learning analysis.

3.5 Recommendation System

The recommendation system employs both collaborative and content-based filtering techniques to suggest personalized learning content to users.

3.5.1 Collaborative Filtering

- User-user similarity metrics, such as cosine similarity, are calculated based on user profiles or interaction history.

- Similar users are identified, and learning content that they have interacted with but the target user has not is recommended.

3.5.2 Content-Based Filtering

- The similarity between the user's tags and the content representations is computed using cosine similarity or other distance metrics.

- Learning materials that closely match the user's interests, preferences, and learning goals are recommended.

3.5.3 Cosine Similarity

Cosine similarity is a technique used in machine learning to measure the similarity between two or more vectors. It's often used to measure document similarity in text analysis.

4. PERSONALIZED LEARNING SYSTEM DESIGN

4.1. Data Collection and Preprocessing

In this stage, data is collected from various sources, including user profiles, learning materials, assessments, and interaction logs. The collected data is cleaned, transformed, and normalized for further analysis.

Table -1: Sample Data Collection

Data Source	Data Points
User Profiles	User ID, Learning Preferences, Historical Performance
Learning Materials	Content ID, Tags, Metadata, Difficulty Levels
Assessments	Assessment Results, Question Bank
Interaction Logs	User Actions, Time Spent, Frequency of Access

4.2. Feature Extraction with TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) is used to convert tags into a numerical representation. TF-IDF indicates how significant a tag is within the dataset. Here's an example of the tags extracted from the learning content and their TF-IDF values:

Table -2: Sample Content wise Tag List

Content	Tags (Extracted)
Content 1	JavaScript, Web Development, Frontend Development, Backend Development, Career Development
Content 2	JavaScript, Web Development, History, Evolution, Programming Languages
Content 3	JavaScript, Programming Basics, Variables, Data Types
Content 4	JavaScript, Programming Basics, Control Flow, Loops

Using TF-IDF, we generate a matrix representing the relevance of each tag within the dataset:

Table -3: TFID Vectors

Tag	Doc-1	Doc-2	Doc-3	Doc-4
JavaScript	0.5	0.6	0.7	0.8
Web Development	0.3	0.2	0.1	0.0
Programming Basics	0.4	0.5	0.6	0.7
Variables	0.1	0.2	0.3	0.4
Data Types	0.2	0.3	0.4	0.5

4.3. Personalized Recommendations with Cosine Similarity

Cosine similarity helps calculate the degree of similarity between user-interested content and other content in the database. It measures the angle between two vectors, indicating their closeness. Here's an example of cosine similarity scores between user-interested content and other content:

Table -4: Similarity scores of Different Documents

Document	Cosine Similarity Score
Content 1	0.3
Content 2	0.5
Content 3	1.0
Content 4	0.6

The similarity scores guide personalized recommendations, suggesting relevant content to users based on their interests.

4.4 Adaptive Learning Paths

Scores Adaptive learning paths adjust to each user's progress and performance, using data on user interactions, assessments, and time spent on learning activities. These paths change as users advance, offering personalized routes through the learning process.

Here's how adaptive learning paths might look for different users based on their progress:

Table -5: Adaptive Learning Paths

User ID	Progress	Adaptive Learning Path
101	50%	JavaScript Basics -> Functions -> Loops
102	80%	Data Types -> Variables -> Scope
103	30%	Web Development Basics -> JavaScript Basics

5. CONCLUSION

This paper's personalized learning system redefines education by offering customized learning paths, adaptable content, and personalized recommendations. Through user profiles, metadata from learning content, and machine learning models like cosine similarity, the system creates an individualized learning environment tailored to user needs.

The system's database design ensures efficient storage and retrieval of user data and learning content. By employing Term Frequency-Inverse Document Frequency (TF-IDF) and cosine similarity, the system accurately recommends relevant content, guiding users toward resources that align with their learning goals and interests.

Adaptive learning paths enable users to progress at their own pace, providing flexibility while maintaining a level of challenge and engagement. The scalable and modular architecture supports various use cases, from educational institutions to corporate training programs.

In conclusion, this personalized learning system represents a significant step in educational technology. By adapting to individual learning styles and harnessing data-driven insights, the system delivers a learning experience that is both engaging and effective. Ongoing data analysis and user feedback ensure the system continues to evolve, meeting the personalized learning needs of its users.

REFERENCES

[1] Bernacki, Matthew & Greene, Meghan & Lobczowski, Nikki. (2021). A Systematic Review of Research on Personalized Learning: Personalized by Whom, to What, How, and for What Purpose(s)?. Educational Psychology Review. 33. 1-41. 10.1007/s10648-021-09615-8.

[2] Akavova, Aida & Temirkhanova, Zarema & Lorsanova, Zarina. (2023). Adaptive learning and artificial intelligence in the educational space. E3S Web of Conferences. 451. 06011. 10.1051/e3sconf/202345106011.

[3] A. R. Lahitani, A. E. Permanasari and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 2016 4th International Conference on Cyber and IT Service Management, Bandung, Indonesia, 2016, pp. 1-6, doi: 10.1109/CITSM.2016.7577578.

[4] Roy, D., Dutta, M. A systematic review and research perspective on recommender systems. J Big Data 9, 59 (2022). https://doi.org/10.1186/s40537-022-00592-5