NavaJyoti: An International Multi-Disciplinary Bi-annual Online Journal
Volume 7, Issue 1, August 2022

Prediction of Individual Educational Materials for Students by Data Mining Techniques

Indu Joseph Thoppil [#1] and K. Ashtalakshmi [*2]

[1,2] *St. Francis De Sales College, Bangalore*

indujoseph@sfscollege.in
kashtalkshmi@sfscollege.in

***Abstract*: The use of machine learning algorithms in education has gained significant attention in recent years. These algorithms can predict academic activities, such as student learning preferences and academic performances, which can help educational institutions identify weaker students early and develop strategies to improve their learning outcomes. By combining information about students' learning styles, objectives, and interests, the system can personalize its content to suit individual needs. For example, if a student struggles with quadratic and linear equations, the instructor can use machine learning algorithms to identify the student's preferred learning methods and create a strategy that fills in any knowledge gaps. However, machine learning algorithms have limitations, including the potential for errors, data acquisition challenges, and time-consuming issues. Therefore, it is crucial to carefully design, test, and refine these algorithms to ensure their accuracy and effectiveness in forecasting academic events. Despite its limitations, machine learning remains a promising technology that can help educational institutions make informed decisions to improve their students' learning outcomes.**

***Keywords*: Educational Datamining, Machine Learning Algorithms, Recommender Systems**

1. INTRODUCTION

Advances in information technologies and the growth of educational media have created new opportunities for using data and analytics in education, teaching, and learning. In addition to their use in formal education, these tools offer potential benefits to professionals outside of education as well. Instruction is a crucial component of any educational endeavor. Effective instruction can help students grasp new concepts, develop analysis skills, and gain and enhance the understanding of the subject of thought. In an online environment, it is important to create engaging and interactive instruction materials that cater to the diverse learning styles of students.

Written e-books can be an effective instructional tool as they allow students to read at their own pace and review the material as many times as needed. Video lectures can provide a more dynamic learning experience and can help to convey complex concepts more easily. Interactive lessons, such as quizzes, games, and simulations, can help students to apply their knowledge and reinforce their understanding of the subject matter.

Instruction should be designed with clear learning objectives in mind, and assessments should be aligned with these objectives to measure students' progress effectively. Effective instruction should also be accessible to all students, regardless of their backgrounds or abilities.

The teaching materials must actively include the pupils, no matter what form they take. In a physical classroom, it's simple to see whether students are paying attention while watching the presentation, taking notes, or displaying other attentive behaviors. But this isn't always the case online. Since we can't see our students when we are teaching them online, it's crucial to organize your materials in a way that will hold their interest. While the results of your assessments will ultimately reveal how well students paid attention to the course material, by designing the materials to maximize their chances of success, we increase the likelihood that they will put in the necessary effort to successfully master their learning objectives.

Traditional educational approaches have frequently relied on a one-size-fits-all approach to instructional materials and exams. However, technological advancements, particularly in machine learning, have enabled the creation of personalized learning experiences that can improve student engagement and retention. Machine learning algorithms can process massive amounts of data, such as a student's learning patterns, performance, and preferences. The system can generate personalised recommendations for learning materials, assessments, and even teaching styles based on this data. Personalized learning can increase student engagement and motivation by providing content that is tailored to their specific needs and interests. It can also help educators better understand the strengths and weaknesses of their students, allowing them to provide more effective support and guidance.

Learning process using AI [6] in education is still a relatively new pitch, with numerous challenges and potential drawbacks to consider. Concerns have been raised about data privacy and security, as well as the possibility that algorithms will perpetuate biases or reinforce existing inequalities. Overall, while machine learning has the potential to transform education and provide personalized learning experiences for all students, its implementation must be approached with caution and careful consideration of the potential risks and benefits

These days, technology is so essential for student engagement because Gen Z is so reliant on the Internet and their smartphones. 69% find it difficult to go more than 8 hours without using the Internet. If that doesn't say anything, consider the 27% of people who can't operate for more than an hour without the Internet.With the help of ML-based AI infused smart tutors, educational institutions can accomplish two goals at once: retain students in the learning process and assist coaches and teachers in varying their teaching approaches. Many different tools can be employed as intelligent tutors that lead a student through a customized learning route and keep the audience interested and motivated.

Furthermore, the items to be recommended in the educational recommendation system are frequently related to educational resources such as books, articles, lectures, and courses. To recommend the most appropriate educational resources, the system must consider the users' educational level, learning objectives, and interests.

Another distinction is that educational recommendation systems frequently involve personalized learning. This necessitates the system tracking the user's learning history and providing recommendations tailored to their specific needs and learning style.

Furthermore, social features such as collaborative learning and peer-to-peer recommendations may be incorporated into the educational recommendation system. Users can now share and recommend educational resources to one another, resulting in a more diverse and comprehensive set of recommendations.

Over the years, recommendation systems have become increasingly important in e-commerce, social media, and content distribution platforms. The idea behind a recommendation system is to predict a user's preference for a particular item based on their past behavior, preferences, and the behavior of similar users.

One of the earliest examples of a recommendation system is the content-based filtering approach, which relies on a user's past behavior to recommend similar items. For instance, if a user has rated several action movies highly, the system may recommend other action movies based on their similarity in genre, director, or actors. [9]

2. LITERATURE REVIEW OF RECOMMENDATION SYSTEMS

The educational recommendation system differs from the general recommendation system. Here we use the ranking system, the emphasis on personalized learning, and the inclusion of social features.

A correct formalization of the predictive problem in recommender systems. Given a set of users U and a set of items I, the goal is to predict the rating or preference that a user u in U would give to an item i in I that they have not yet interacted with. This can be represented mathematically as follows:

Let R be the user-item rating matrix where R_{ui} is the rating given by user u to item i. For a user u and an item i that they have not interacted with, the goal is to predict R_{ui}.

A recommender system seeks to predict the usefulness of an item for a given user. This is achieved by analyzing the user's profile, the features of each item, and the user's previous interactions with similar items. Predicted utility can be used to generate a list of items for users to consider.

Recommender Systems are used widely in a range of applications, including e-commerce, social networks, and online advertising. By making personalized recommendations, they hope to improve the user experience and increase engagement.

One common approach to this problem is to use collaborative filtering, which leverages the ratings of other users to make predictions. Specifically, given a target user u, the algorithm finds other users who have rated similar items to the item in question i, and uses their ratings to predict the rating that u would give to i.

Another approach is content-based filtering, which relies on the features of the items and the preferences of the user to make predictions. In this case, the algorithm would use the known

preferences of the user and the features of the item to predict the rating that the user would give to the item.

There are also hybrid approaches that combine collaborative filtering and content-based filtering to improve the accuracy of predictions. Hybrid approaches improve recommendation accuracy and coverage. Furthermore, knowledge-based techniques make recommendations based on domain knowledge and user feedback, and online learning techniques can be used to continuously adapt and improve the recommendation system over time.

3. DIFFERENT APPROACHES FOR RECOMMENDER SYSTEM

3. 1. CONTENT-BASED APPROACH

The content-based recommendation systems are based on the principles of information retrieval and filtering, which originated in the field of information science. This method entails analyzing[1] the properties of items in a database (such as text, keywords, or other features) and then matching those properties to the preferences or interests of individual users.

Traditional information retrieval systems differ from content-based recommendation systems in that the latter are personalized to individual users based on their profiles. User profiles contain information about their previous behavior, preferences, and interests, which is used to tailor recommendations to their specific needs. These user profiles can be created explicitly, through surveys or questionnaires, or they can be inferred implicitly from user behaviors on their social medias and other websites.

Content-based filtering, relies on the features of the items and the preferences of the user to make predictions. In the case of text-based items, such as web pages, the content is usually described with keywords or other textual features.

The example you mentioned, the Fab system, represents the content of web pages using the 100 most important words. This means that the system identifies the most significant words in the content of each web page and uses them as the basis for recommending other similar pages to users.

Content-based recommendation systems can be useful when users have well-defined preferences for specific features or characteristics of items, such as specific keywords or topics. However, they may not be as effective when users have more diverse or complex preferences, or when the content of items is not easily described with keywords or other textual features.

TF-IDF stands for "Term Frequency-Inverse Document Frequency". It is a statistical measure that is used to evaluate the relevance of a term (or word) in a document within a collection of documents or corpus. The metric is commonly used in information retrieval and text mining applications.

The TF-IDF metric consists of two parts: Term Frequency (TF) and Inverse Document Frequency (IDF).

Term Frequency (TF): This measures the frequency of a term in a document. It is calculated by dividing the number of times a term appears in a document by the total number of terms in that document.

TF(t,d) = (Number of times term t appears in document d) / (Total number of terms in document d).

Inverse Document Frequency (IDF): This measures the importance of a term in a collection of documents. It is calculated by dividing the total number of documents in the corpus by the number of documents that contain the term, and then taking the logarithm of that ratio.

IDF(t,D) = log_e(Total number of documents / Number of documents with term t in it)

The TF-IDF metric is then calculated as the product of TF and IDF:

TF-IDF(t,d,D) = TF(t,d) x IDF(t,D)

The resulting TF-IDF score for a term t in a document d and a corpus D provides a measure of how important the term is in the document relative to the other documents in the corpus. Terms with high TF-IDF scores are considered to be more relevant to the document, while terms with low scores are less relevant.

The TF-IDF measurement combines these two components by multiplying her TF and IDF values for each keyword. The result is a weighted score that reflects the importance of the keyword within the document or corpus.

TF-IDF is commonly used in information retrieval applications such as search engines, document classification, and clustering. It is also used in natural language processing tasks such as text summarization, topic modeling, and sentiment analysis.

Some examples of machine learning techniques used in recommendation systems include:

Collaborative filtering: This technique analyzes user behavior to find similarities between users and recommend items based on what similar users liked.

Content-based filtering: It relies on the features of the items and the preferences of the user to make predictions.

Matrix factorization: This technique factorizes the user-item matrix into low-dimensional representations to identify hidden patterns and relationships.

Deep learning: This technique uses neural networks to learn complex relationships between items and users and make recommendations based on that learning.

Bayesian classifiers are a type of probabilistic classifiers that use Bayes' theorem to classify data based on a set of features or attributes. They are often used in machine learning and natural language processing applications, particularly in text classification.A Bayesian classifier can be trained on a set of labeled data (e.g., user ratings of items) to predict whether a new item will be liked or not liked by a user. The naive Bayesian classifier assumes that the features of an item are independent of each other, which can simplify the modeling process.

Video analysis focuses on extracting metadata from recorded video footages. These data can be used in applications like searching, categorization, summarization and event recognition. The process tends to transform audio and images into meaningful components.

The growing use of educational software and online learning platforms is generating vast amounts of data on student interactions, performance, and behavior. You can analyze this data to understand how your students are learning and identify areas for improvement.

Universities are increasingly turning to analytics to make more informed decisions about the services they offer their students. By analyzing data about student performance and behavior, educational institutions can identify patterns and trends that can help design more effective teaching and learning strategies to pinpoint students who are at risk of dropping out or need additional support. In addition to improving grades and student retention, analytics can also be used to improve operational efficiency and identify areas where resources can be allocated more effectively. For example, institutions can use data to identify courses that are popular with students, allocate resources to areas of high demand, and identify areas where course content and delivery can be improved.

Overall, the use of analytics in educational field will revolutionize the way institutions work thus improving the student outcomes. However, it is important to ensure that data is collected and used ethically and student privacy is protected.

3.2 USER-BASED COLLOBORATIVE FILTERING TECHNIQUE

User-based collaborative filtering is a technique used in recommender systems to recommend items to users based on the preferences of other users who are similar to them.

The approach is based on the assumption that users who have similar preferences in the past are likely to have similar preferences in the future. The technique works by finding users who have rated items similarly to a target user and then recommending items that the similar users have liked but the target user has not yet rated.

The following steps are typically used in user-based collaborative filtering:

Create a user-item matrix: This is a matrix where the rows represent users and the columns represent items. Each cell in the matrix represents the rating that a user has given to an item.

Calculate user similarities: Calculate the similarity between each pair of users based on their ratings. Common similarity metrics include cosine similarity, Pearson correlation, and Jaccard similarity.

Find similar users: Identify a set of similar users to the target user based on their similarity scores.

Generate recommendations: Recommend items to the target user based on the items that the similar users have liked but the target user has not yet rated. This can be done by calculating the weighted average of the ratings of the similar users for each item and recommending the items with the highest weighted average.

Many studies have shown that this method is effective, and it has been used in recommender systems for online learning to personalize learning by generating relevant items for targeted students based on their activities and ratings.

3.3 CONTEXT-SENSITIVE RECOMMENDER SYSTEMS

Context-sensitive recommender systems are a type of recommender system that take into account contextual information in addition to the user's historical preferences and item characteristics. Contextual factors may include time of day, location, weather, social network information, and activity information. [7].

For example, context-sensitive music recommender system may take into account the user's location, time of day, and listening history to suggest music that is appropriate for the current context. If the user is at the gym in the morning, the system may suggest upbeat workout music, while in the evening at home, it may suggest more relaxing or mellow music.

Context-sensitive recommender systems can improve the experience of the user by providing more relevant and personalized recommendations, which can increase engagement and satisfaction. They can also be beneficial for businesses by increasing sales and customer loyalty.

4.SUGGESTED FRAMEWORK

Below are possible frameworks for predictive engines for educational data mining.[2],[3],[4].

Data Collection: We collect data from various sources, including student information systems, learning management systems, online learning platforms, and student feedback surveys.

Data preprocessing: Analysis can be done after data cleansing. This step involves identifying and addressing missing data, outliers, and data quality issues.

Data analysis: In this step, different data mining techniques can be applied to the preprocessed data, such as clustering, classification, and association rule mining. The goal is to identify patterns, trends, and relationships in the data that can help understand student learning behaviors, identify factors that impact student performance, and develop personalized interventions.

Model selection: Select an appropriate predictive model based on the nature of the problem and available data. Following are the list of models used in educational data mining a) Decision Tree b) Logistic Regression c) Neural Networks.

Train Model: Uses preprocessed data to train a model of your choice.

Model evaluation: This step involves comparing the results of the analysis with the ground truth, such as teacher feedback or assessment scores, and assessing the impact of the interventions on student learning.

Model Optimization: Optimize selected models to improve performance. This includes tuning hyperparameters, changing model architecture, or using ensemble techniques.

Deploy: Deploy optimized models into real-world environments such as learning management systems and online platforms.

Monitoring: Continuously monitor the performance of deployed models and update as needed based on new data and feedback.

5. CONCLUSION

The use of machine learning techniques in student academic growth analysis and recommendation of instruction materials can greatly benefit educators and students alike. By analyzing student data, such as their performance on assessments, attendance, and engagement in the classroom, machine learning algorithms can identify patterns and predict future academic growth. This information can then be used to recommend personalized instruction materials to students, based on their individual needs and learning styles.

The use of binomial logical regression, decision trees, entropy, and KNN classifiers are common machine learning techniques used in educational data mining. These algorithms can be trained on historical student data and used to make predictions about future student performance. By using these predictions, instructors can provide more targeted and effective instruction, which can improve student learning outcomes.

In the future, additional features can be added to the dataset to improve accuracy in predicting student academic growth. For example, data on student engagement with online learning platforms or social network interactions can be incorporated into the analysis. This can provide a more comprehensive understanding of the student's learning behaviors and enable more accurate predictions of future academic growth.

Overall, machine learning has the potential to greatly improve the educational experience for both instructors and students. By leveraging the power of data and algorithms, educators can provide more personalized and effective instruction, which can lead to improved academic outcomes.

This framework can be used to build predictive engines that provide insight into student performance, identify students at risk, and suggest personalized interventions to improve learning outcomes.

6.REFERENCES

[1] M. Allahyari, S. Pouriyeh, M. Assefi, S. Safaei, E. Trippe, J. Gutierrez, and K. Kochut. 2017. A Brief Survey of Text Mining: Classification, Clustering and Extraction Techniques. ArXiv (07 2017).

[2] P. Crespo and C. Antunes. 2012. Social Networks Analysis for Quantifying Students' Performance in Teamwork. In Proceedings of the 5th International Conference on Educational Data Mining. New York, NY, USA, 232–233.

[3] S. Das. 2010. News Analytics: Framework, Techniques and Metrics. The Handbook of News Analytics in Finance (03 2010). https://doi.org/10.2139/ssrn.1814258.

[4] T. Hassan, B. Edmison, L. Cox, M. Louvet, D. Williams, and D.S. McCrickard. 2020.Depth of Use: An Empirical Framework to Help Faculty Gauge the Relative Impact of Learning Management System Tools. In 2020 ACM Conference on Innovation and Technology in Computer Science Education (USA) (ITiCSE '20). Association for Computing Machinery, 47–53. https://doi.org/DOI:https://doi.org/10.1145/

3341525.3387375.

[5] S. Kotsiantis, K. Patriarcheas, and M. Xenos. 2010. A Combinational Incremental Ensemble of Classifiers as a Technique for Predicting Students' Performance in Distance Education. Knowledge-Based Systems 23, 6 (2010), 529–535. https://doi.org/10.1016/j.knosys.2010.03.010.

[6] S. B. Kotsiantis. 2012. Use of Machine Learning Techniques for Educational Proposes: A Decision Support System for Forecasting Students' Grades. Artificial Intelligence Review 37, 4 (01 Apr 2012), 331–344. https://doi.org/10.1007/s10462-011-9234-x.

[7] V. Ramasamy, J. D. Kiper, O. Hemraj, and U. Desai. 2019. Analyzing Link Dynamics in Student Collaboration Networks using Canvas - A Student-centered Learning Perspective. In IEEE Frontiers in Education Conference (FIE 2019). IEEE, Los Alamitos, CA, USA, 1–9. https://doi.org/10.1109/FIE43999.2019.9028629.

[8] C. Romero, M. Lopez, J. Luna, and S. Ventura. 2013. Predicting Students' Final Performance from Participation in On-line Discussion Forums. Comput. Educ. C (Oct. 2013), 458–472. https://doi.org/10.1016/j.compedu.2013.06.009

[9] C. Romero, S. Ventura, Sebastian, G. Espejo, and C. Martínez. 2008. Data Mining Algorithms to Classify Students. Educational Data Mining 2008 - 1st International Conference on Educational Data Mining (2008), 8–17.

[10] S. Sathya and N. Rajendran. 2015. A Review on Text Mining Techniques. International Journal of Computer Science Trends and Technology (IJCST) (2015),274–284.