International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Contextual Course Clustering: BERT-Enhanced Text Analytics for Personalized Education

Mohd Mustafeez ul Haque[1], Bonthu Kotaiah[2], and Jameel Ahamed[3]

Submitted:10/03/2024 **Revised**: 25/04/2024 **Accepted**: 02/05/2024

Abstract: In the case of online learning platforms, an efficient and effective course recommendation system that provides personalized content is one of the most significant factors in choosing a platform for learning. This research aims to develop a new paradigm incorporating the clustering approach and score-matching technique to guide students in appropriate courses based on their academic records. It uses a stacking technique, namely, bidirectional encoder representations from the transformer (BERT) for topic modeling, singular value decomposition (SVD) for collaborative filtering, term frequency-inverse document frequency (TF-IDF) and word-to-vector (Word2Vec) embeddings for clustering and similarity analysis on the NPTEL and Coursera datasets. These methods are incorporated in the proposed algorithm to recommend courses that are close in academic relatedness to the student's past performance and preferred course offers. Using average similarity metrics in actual performance evaluations shows the ability of the approach to recommend the best courses relevant to each student. The root mean square value and mean absolute error for the BERT-SVD model are 1.3267 and 1.0802, respectively. The findings highlight the potential of advanced clustering and embedding techniques in improving the accuracy and relevance of course recommendation systems in educational settings.

Keywords: *BERTopic Model, Course Recommendation, TF-IDF, Word2Vec, Coursera, Context-aware clustering.*

1. Introduction

In education development, technology-aided learning solutions have evolved into one of the critical strategies for increasing student motivation and satisfaction and improving overall academic performance. In the context of diverse e-learning platforms and the large volume of online faculties and courses, it becomes more critical to provide students with accurate and efficient course recommendations [1]. The rationale for designing recommendation systems for learners' courses is based on the fact that learners' course preferences have changed and are likely to change due to technological growth in current society. Unlike in the past, where students were limited to fixed traditional course selection as each level was designed, there was little consideration of the students' variety of learning styles, preferences, and aspirations. The combined recommendation algorithms for integrated course recommendations take user information like enrollments in previous semesters or courses, academic performance history, and preferences. Due to their ability to provide optimal recommendations for learners' needs and preferences, the systems enhance learner's involvement, satisfaction, and performance levels. Also, people recognize the need for continuous learning to help them learn for their lifetime and equip them with the learning they need to pursue their chosen interests and careers[2]. Course recommendation applications use machine learning and, by considering the user profile, previous learning, and the student's desired career path, provide him with course recommendations. Through an examination of past enrollment statistics, academic results, and user interactions, these systems aim to provide course recommendations that are both timely and valuable, allowing for smoother e-learning experiences and reiterating the benefits of continuous learning. This paper discusses the methodology of the development, methodology of its instruction, and evaluation, intending to produce a deeper understanding of the application of the personalized course recommendation system and its potential future course in the field of education technologies.

The following research proposal based on a novel approach to contextual course clustering is presented in this context: BERT-based topic modeling and SVD with tf-idf and word2vec. Through the utilization of BERT and other current NLP models to achieve progress, the ultimate goal is to transform the process of course recommendations in educational settings.

The research motivation is based on the fact that most clustering algorithms achieve poor performance in clustering educational texts that contain many semantics. Despite the success of such approaches as tf-idf for vector representation and Word2Vec as a neural instantiation of tf-idf representation, they have proven rather inefficient in

[1] *Maulana Azad National Urdu University, Hyderabad, 500032, India*
ORCID ID : 0009-0003-2764-7306
[2] *Central Tribal University of Andhra Pradesh, Vizayanagaram, 535003, india*
[3] *Maulana Azad National Urdu University, Hyderabad, 500032, India*
* *Corresponding Author Email: mohdmustafeezulhaque@gmail.com*

capturing contextual data. Consequently, incorporating popular BERT as a contextualized word embedding model could be a good way to enhance online course recommendations and reduce the limitations mentioned above.

The following are the objectives of this study: To design and assess a new framework for course recommendation in learning, based on BERT-Topic with SVD for personalized recommendations, improved by tf-idf et Word2Vec for benchmarking purposes. Thus, concrete objectives are to compare the ability of BERT to grasp the semantic context, compare the results of various text mining methods for analyzing subsequent variables, and illustrate the advantages of BERT-based models through comparative analysis of the average similarity score. In the end, this work will seek to provide a sound foundation for enhancing and improving adaptive learning systems by providing an innovative structural model that can help to enhance and improve course recommendations individualized based on the needs of each student.

Thus, the proposed architecture results from improving the BERTopic model structure with the CountVectorizer for density calculation, UMAP for visualization, and HDBSCAN for clustering tasks. In addition, SVD is used to predict the score. This allows the identification of similarities and relationships between courses in a more detailed way, as the relationships between a range of courses need to be considered. By extending the contextual information to include the global information in context clustering, the model will provide more personalized context-dependent suggestions for each learner.

Thus, one of this work's important contributions relates to comparing text analytics approaches: BERT, TF-IDF, and Word2Vec. It is thus proposed that by comparing the overall procedures supplemented in this document in terms of average similarity scores, BERT can be shown to excel in identifying the semantic context of educational texts. These results provide empirical evidence of the effectiveness of BERT-based models for course clustering and indicate that enhancing the overall performance of adaptive learning environments may be highly possible.

2. Literature Review

The current state-of-the-art e-learning systems have led to the emergence of novel, promising directions for further work in the context of personalized education and course recommendation. Specifically, Li et al. (2023) proposed a model of deep learning and big data personalized recommendation systems for MOOCs. The same authors have developed an enhanced deep course recommendation system based on the BERT framework and using feature extraction involving enhanced self-attention[3].

This work is built upon the RL algorithm, which was used in the smart e-learning approach proposed by Amin et al. (2023) to enhance the adaptive e-learning process for every learner. Its sequence-based recommendation also modifies the Markov decision process to improve the sequential search for path recommendations. It confronts the query function of selecting appropriate courses that help in the various forms of online learning platforms [4].

The filter recommendation in this study, where Muzdybayeva et al. (2023) focused on course recommendation in higher education, used an algorithm based on matrix factorization. The students' databases include records of the student's past performance and preferences, which are used to suggest likely courses and assist the students in choosing the courses suitable for their career path [5].

In the paper of Jena et al. (2022), a recommender system for an e-learning course inventory using collaborative filtering was developed that includes several techniques, such as KNN, SVD, and NCF. In their work, they used feature subset selection to compare the accuracy of KNN. They found that this model was useful in assisting the learner in arriving at an informed decision as to the course preference most suited to their interest from the available list of courses to enroll in [6].

Semantic coupled course recommendation using deep learning techniques: In the case study conducted by Premalatha et al. (2022), the authors generalized mainly focusing on domain-specific course recommendations. With this, their system can map core course content to predefined domains and train deep learning models to suggest appropriate course offerings based on student's existing domain knowledge, enhancing delivery and completion rates [7].

Guruge et al. (2021) characterized various approaches implemented in CR systems; recent technologies and strategies, including hybrid recommendation and DM methods, are also discussed. As with their findings, they raise possibilities about the current models in the course recommendation, strategies, and research areas that can be useful for the study [8].

Wang et al. (2019) also proposed an attention-enhanced CNN to propose MOOC personalized course recommendations based on user behavior representations and attentional models. It proposed works that offer a more formal method of course recommendation based on course preference that employs deep learning with a basic recommendation process [9].

In particular, Li and Ye (2020) proposed a novel online education model in which course recommendation depends on the collaborative filtering technique. By enhancing the procedures employed in the two traditional collaborative filtering frameworks, their creation provides a dependable course proposition system that is pertinent to the miscellaneous general interests of the particular user and, thereby, ensures more personalized online tuition and learning[10].

Another study by Mondal et al. (2020) suggested designing a course recommendation system, which translates to retaining the score history of learners to provide the learner with an appropriate course that they should take to provide recommendations on the course to be completed by the learner. Their framework uses k-means clustering and collaborative filtering to recommend the course that best suits the individual learner: flexibility [11]. Hence, there is a need to develop aids to assist her in making her own decisions regarding her preferences and preferences for flexibility.

Lin et al. (2021) proposed a new model for making adaptive course recommendations in a MOOC, including dynamic attention and hierarchical RL. This concept of capturing the interest flow of users enhances the relevance of offering courses that meet learners' interests [12].

Based on the above analysis, Morrow et al. (2020) proposed an algorithmic support system for recommending courses and scheduling according to the context of each learner, thus helping students plan a roadmap on how to complete their studies in higher education institutions. Their system also considers students' characteristics and curriculum needs and produces optimal course schedules to improve experience inside and outside the classroom and decrease time to a degree [13].

Kobe Joe & Nago M. (2021) surveyed various methodologies used in course recommender systems, including current trends and methods such as integrating recommendation strategies, data mining methods, and techniques. This is because their work offers an understanding of what currently exists regarding recommendations for course approaches and where future studies could be directed [14].

Li and Kim (2021) developed a deep learning-based course recommendation system with sustainable development in education. High-level user behaviors and coursework features minimize sparsity and scalability issues and foster recommendation solutions. With their DECOR model, they provide strong, accurate course recommendation services that can help increase resource utilization in online education systems [15].

Nguyen et al. (2021) developed a course recommendation model based on learning outcomes. Their framework employs data mining techniques to predict students' future learning outcomes and recommend appropriate courses accordingly. It utilizes collaborative filtering methods to personalize course recommendations, enhancing the learning experience for students and improving academic performance [16].

Another relatively recent work is Wang and Fu (2021), who presented a personalized learning resource recommendation method based on their dynamic, collaborative filtering approach that uses the Pearson correlation coefficient and slope coefficient, among others, for better accuracy and efficiency. Their learning model provides formative learning by suggesting the latest and most suitable literature the learner wants to read [17].

Given the developments in the frameworks detailed in this literature review, employing an improved topic modeling framework for the recommendation of personalized courses based on BERT combined with tf-idf and Words2Vec for comparative analysis will considerably improve the precision and efficiency of the developed frameworks. The context course clustering architecture currently incorporates BERTopic, CountVector, UMAP, and HDBSCAN to explore deeper course similarity, supporting the targeted learning process of a single-user system. It is believed that incorporating BERT-based models will outcompete conventional recommendation techniques, including collaborative filtering and deep reinforcement learning models; this will lead to better results involving improved average similarity scores and general performance of the recommender system. This hypothesis postulates that the given architecture of contextual course clustering will greatly advance the implementation of adaptive learning systems. It should help users obtain better and more relevant courses suggested for improving their literacy, learning enhancement, and experiences.

While a substantial body of literature extends across multiple domains to personalized course recommendation systems involving deep learning methodologies, collaborative filtering techniques, and the adoption of semantic analysis frameworks, it is critical to note that there is sparse analysis of BERT-based models in this context. Although there are studies that discuss the use of BERT in recommendation systems, few studies have investigated the use of BERT in personal course recommendation systems. Moreover, the methods and frameworks used in the literature are independent, with relatively few comparative studies comparing one model or technique to another. Thus, it may be concluded that the research gap is based on the absence of a comprehensive analysis of BERT-enhanced text analytics in the context of the proposed personalized course recommendations and comparison with other methods, which would allow for determining the effectiveness of the offered approach, as well as its advantage over others in terms of increasing the accuracy of recommending items and user satisfaction.

In this study, we use a BERTopic model and compare it to the TF-IDF and Word2Vec models for recommending courses personalized to each individual. The next part of the study will explain how we conducted our research.

3. Methodology

This research systematically surveys and compares the course suggestion models used in online education systems to understand their usefulness. The process includes data acquisition, cleaning and wrangling, model development, and experimental assessment. First, enrollment information concerning students' success and academic records from several academic sessions are collected and stored in a single format. Next, different course suggestion models are applied, such as BERTopic clustering, TF-IDF vector, agglomerative clustering, Word2Vec, and K-Means clustering. The evaluation of each model is performed according to how effective it is in recommending the most appropriate course for the students by measuring values such as the average of the similarity scores and the accuracy of the predictions. This section presents a systematically designed approach that was followed during this study and enabled the gathering salient information on the strategies for course recommendations in online environments.

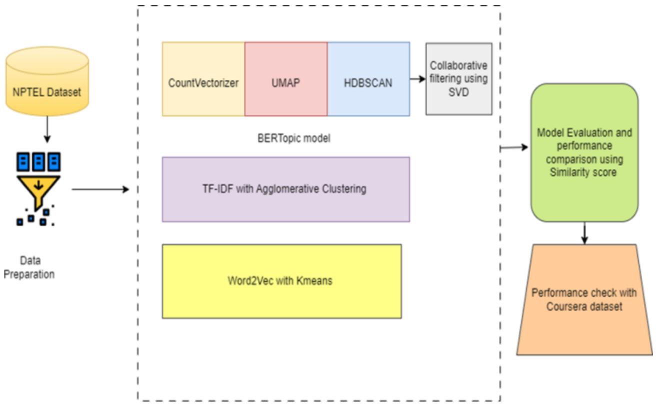

Fig 1: System architecture

3.1 Dataset

National Programme on Technology Enhanced Learning (Nptel)course dataset: In this research study, data collection and preparation included obtaining enrollment information and the results of various academic year examinations. Each of these data sources will help determine students' future performance rates and enrollment. Similarly, we utilized the Nptel Course dataset.

Coursera dataset consists of information about the course, such as the title, description, instructor's name, number of students who have enrolled in the course, ratings, and reviews. It provides knowledge on the current trends in online education and, therefore, can be helpful for market research, the design of recommendation systems, and educational research. Investigators investigate features such as course enrollment, learner participation, and achievement, and educators seek course features that existing MOOC platforms lack or have preferences for. By forecasting enrollment and completion rates, actionable insights can be used to strengthen platform strategy predictive modeling. The dataset has provided useful insight into the online learning landscape and ways to advance learning models. In this research, the Coursera dataset is used to compare model performance.

3.2 Data Preparation

First, the concerned department's enrollment and exam results in Excel files and respective academic years were derived from the Nptel dataset. These files included student ID numbers, courses offered, the lectures/sections they attended, and their respective test performances. Every file was organized by year, making it possible to trace student performance changes.

After obtaining the Excel files, the files are transformed to compile all the DataFrames into a single DataFrame. This consolidation was necessary to conduct the data analysis quickly and efficiently and more easily gain an overall picture of the student's academic progress.

The consolidation process involved several key steps:

1. First, each Excel file was cleaned so that all necessary data were fetched, and the information included the names of the students, the course of the learner, and the examination results. In this process, the work performed with the help of the necessary Python libraries, such as Pandas and Open Python, was severe.
2. Data Integration: While working on each file, the data were combined as a data type called DataFrame in one set. This integration entailed the following: It enabled all the information that should be collected to be obtained in one form, thus simplifying the process of analysis and interpretation.
3. Data Mapping: To combine the examination results with the enrollment data, the two files were merged based on key fields, which included student names and course codes. These mappings were useful in correlating exam results with the courses taken by each student through the students' and course performance analysis.
4. Data Cleaning: Disparate entries within the DataFrames were also processed before the DataFrames were merged to create the consolidated DataFrame. At other times, it was about managing nulls in the datasets, removing the redundant records from the databases, or standardizing the units of measurement of various data fields to enhance the accuracy of the data as and when available.

5. Data Validation: Validation of self-checking and self-confirmation exercises was performed since they confirmed the correct consolidation of the data. This process involved verifying the exercise of the student's association with the courses and then comparing the scores obtained from the examination records with the scores obtained from the enrollment data records.

This case created a detailed and structurally acceptable data frame with various details on enrollments and their corresponding exam results for multiple academic periods. To draw inferences and conclusions about students' performance and enrollment, this compound dataset was used for further analysis.

3.3 BERTopic model

In line with the methodologies outlined, the process following data collection entails data preparation through BERTopic Modeling, which aims at clustering sequences to determine their existing relationships. The purpose of the process is to collect data from semantically similar courses together to provide a better understanding of the curriculum and to assist in the course recommendation system.

We used BERTopic modeling, a named topic model based on BERT embeddings that seek to capture semantic similarity between different sentences or, in our case, course names. In particular, BERT, among the most recent and effective methods for language representation, supplies enhanced contextualized embeddings that represent the given text much more accurately in terms of its semantic meaning than more archaic methods [18].

The first step used the BERTopic library to identify the topics within the course names. This entailed mapping each course name to its BERT vector and then segregating the identified vectors into coherent topics using clustering techniques. Thus, it would be easier to extract further queries or find various topics at the core of the course dataset to discover inherent semantic associations between the courses.

Since BERT embeddings produce high-dimensional vectors, tools for reducing dimensions were used to make the datasets more analyzed and visually presentable. As a dimensionality reduction technique, we used UMAP, a nonlinear approach known for its capacity to maintain high-dimensional data's local and global spatial properties. UMAP reduces the BERT embeddings to a much lower dimension while retaining the relations to other courses, thus allowing for the visualization and interpretation of data with complicated associations.

3.3.1 CountVectorizer

This method converts a collection of text documents into count vectors, giving each token frequency in the text, which is a basic yet powerful way to analyze text. The given course title is split into a set of technically independent words; each word's respective count is determined. These counts serve as the beginning of the feature vectors used for every course. This approach provides a foundation for downstream tasks such as clustering, where one can numerically measure the similarity between two-course titles by comparing their corresponding word frequency vectors.

3.3.2 Uniform manifold approximation and projection

Many data points involve complicated relationships to other courses, and UMAP effectively decreases the dimensionality of BERT embeddings to provide visualizations and interpretations of such data.

UMAP preserves the local and global structure of the data by minimizing the following objective function:

$$\min_{y} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} \left(D_{ij} - \log\left(\frac{1}{n}\sum_{k=1}^{n} e^{-D_{ik}}\right) - \log\left(\frac{1}{n}\sum_{k=1}^{n} e^{-D_{jk}}\right) + \log\left(\sum_{k=1}^{n} e^{-D_{ij}}\right)\right) \quad (1)$$

where:

- y is the low-dimensional embedding of the data,
- D is the pairwise distance between data points *i* and *j* in the high-dimensional space,
- w_{ij} is a weight function,
- n is the number of data points.

3.4 Hierarchical density-based spatial clustering of applications with noise (HDBSCAN) algorithm

Once the dimensionality of the BERT embeddings was reduced, we applied the hierarchical density-based spatial clustering of applications with noise (HDBSCAN) algorithm for clustering. HDBSCAN is a density-based clustering algorithm capable of effectively identifying clusters of varying shapes and sizes while handling noise. By employing HDBSCAN to reduced-dimensional BERT embeddings, we aimed to partition the courses into clusters based on their semantic similarity, thereby revealing cohesive groups of courses sharing common themes or topics.

The probability density of a data point x at a distance r from its nearest neighbor is estimated using the following formula:

$$Density(x, r) = \frac{k}{V(r)} \quad (2)$$

Where k is the number of points within distance r of x, and $V(r)$ is the region's volume enclosing x within distance r.

HDBSCAN consequently builds a hierarchical version of the clusters by merging clusters to ensure that the variations in density are maximized in the resulting clustering pattern.

The proposed method can extract semantically driven associations in a dataset containing course data by combining BERTopic modeling, dimension reduction, and HDBSCAN clustering in a context-aware manner. Thus, using natural language processing and machine learning algorithms will further improve the understanding of the curriculum and improve course recommendation systems in online education. This methodology defines course semantics explicitly and paves the way for future research toward efficient education data mining and recommendations.

3.5 Collaborative filtering using SVD

Before applying singular value decomposition (SVD) in collaborative filtering, student performance on the course is normalized to the range between 1 and 10, and the right average is employed if a value is missing. Here, the student list is constructed from the students' names, the courses are taken, and the normalized mark list is employed for training the SVD model. This work also shows how the model explores every student's behavior with the course and makes the right rating prediction, which is equally helpful in identifying the student rating with the course based on the performance.

The best approach to collaborative filtering has been established as using singular value decomposition on the data where the data is normalized in a student course interaction matrix form. Singular value decomposition (SVD) is then performed on this matrix to decompose it into several components that best represent the students and the courses. The model then uses these predictors to predict student ratings, which results in a better understanding of how student features influence course properties. The main aspects of this method involve recommending courses based on matrix factorization and predicting missing rating values to make educational courses more personalized.

3.6 TF-IDF with Agglomerative Clustering

In this section, we elaborate on another method discussed in the previous section: course suggestion through term frequency-inverse document frequency vectorization with agglomerative clustering. This leads to offering ideas on a logical sequence of courses whose mapping can easily be done based on students' characteristics. Here, we expound more on the details of this approach and dive deeper into its intricacies.

TF-IDF vectorization, where the data in a text format is converted to a vectorial representation and is optimal when determining the relative importance of a given word in a collection of documents, is one of the most common textual feature extraction techniques in natural language processing [19]. The TF-IDF score for a term t in document d within a corpus is calculated as follows:

$$\text{TF-IDF}(t,d)=\text{TF}(t,d)\times\text{IDF}(t) \quad (3)$$

Where:

- $\text{TF}(t,d)$ represents the term frequency of term t in document d, indicating how frequently the term appears in the document.
- $\text{IDF}(t)$ is the inverse document frequency of term t calculated logwhere N is the total number of documents in the corpus, and nt is the number of documents containing the term t.

Using the TF-IDF scores, we can calculate every term in the document and obtain a representation of the terms in repeated vectors that vary by the term's usage in the document and throughout the corpus.

After the course names have been changed in the TF-IDF vectors, the next step is to cluster them to enable group analysis of similar courses. We chose agglomerative clustering, a hybrid of a hierarchical clustering approach in which the clusters change over time through merger and fusion processes, with the linkage criterion guiding the process.

The agglomerative clustering procedure is as follows: The initial condition assigns each candidate course a cluster of its own. In the merge step, the algorithm joins the nearest clusters and transforms them into a single cluster until the required number of clusters is reached. A metric such as the cosine of the angle between the vectors and adjacent clusters defines the distance between clusters. The next criterion is the linkage criterion, which specifies how the distances between the circles representing clusters are computed, and the choices include 'complete', 'single', and 'average'. The algorithm runs iteratively, meaning the next step is performed until a certain condition is met. These conditions may include a fixed number of clusters or when the distance between the clusters is greater than a predetermined acceptable limit.

3.7 Word2vec with K-Means

In word2vec with the K-means approach, access to word2vec was employed to obtain the semantic similarity measures of the names of the courses. Word2vec is a famous model in natural language processing (NLP), where it works by learning the vector space representation of a word by capturing its semantic relation with the other word given by its co-occurrence pattern [20].

First, we identified the corresponding course name in the array to solve the problem of obtaining the corresponding course name for each short string more quickly by training a Word2Vec model on a corpus of course names. To do so, the names of the courses were passed to the Word2Vec algorithm combined with the text corpus data and used to generate high-dimensional vector values for the words. In line with what has been previously stated, it has been posited that it is possible to generate dense vector representations in the format of word2vec, using the text in which the words in a set of words are in the name of a set of courses; the text contains semantic information concerning the courses.

After training the Word2Vec model, we proceeded with K-means clustering of the extracted Word2Vec vectors. K-Means is an out-of-bend procedure of clustering that forms k clusters of samples closest to the points or mean of that cluster. In this application, the input features used to cluster the data were word2vec embeddings of course names.

K-Means successively groups the Word2Vec embedding by assigning each point the nearest centroid and then recalculates the centroids with the points' radius to reduce the sum of the squared distances of the members in a cluster. This leads to k clusters, in which each cluster represents several names of the course that, based on the Word2Vec concept, possess semantically related meanings.

Once the program reaches the clustering stage, the next step is to find the nearest course most relevant to the student in each cluster obtained. This involves using a formula to determine the arithmetic mean of all the points in each cluster and then calculating the square root of the sum of squares of the difference between the student profile, a vector, and the center of each cluster. Specifically, the course of study closest to that particular cluster's center point is recommended to the student as the most suitable course in that semantic grouping.

Therefore, while the improvement in combining Word2Vec embeddings with K-Means clustering is a positive effect of this approach, it enhances the potential of given solutions in recommending the course by comparing their semantic similarity; thus, the recommendations made would be accurate and would meet the student's interest.

Model evaluation and performance comparison using the similarity score

In assessing the effectiveness of singular value decomposition (SVD) for collaborative filtering, two primary metrics are utilized: the root mean square error (RMSE) and mean absolute error (MAE), which are the types of performance metrics that measure the difference between the predicted and actual values. The RMSE is calculated as the square root of the mean square error of the estimated and actual rating values, and the closer the value is to zero, the higher the accuracy of the model. The same is true for the MAE, which seeks to find the best prediction error through the average of the absolute differences.

Validation involves applying the model to a completely different data set to compute the RMSE and MAE and evaluate the model's efficiency in accurately predicting unseen rating values. Cross-validation is applied to check the model's stability using multiple subsets to increase its validity. The RMSE and MAE indicate the model's accuracy and are useful for recommending related courses based on students' historical interactions. These strict evaluation checks guarantee that the process utilized by SVD achieves its primary goal of employing collaborative filtering to successfully pinpoint courses the student will likely take based on their preference.

$$RMSE = \sqrt{\frac{1}{N}\sum_{(u,i)\in TestSet}(\hat{r}_{ui} - r_{ui})^2} \quad (4)$$

$$MAE = \frac{1}{N}\sum_{(u,i)\in TestSet}|\hat{r}_{ui} - r_{ui}|) \quad (5)$$

where N is the total number of user-course pairs in the test set.

$\hat{r}_{ui}$ Predicted rating by the SVD model for user u on course i.

r_{ui} Actual rating is given by user u for a course i.

To determine the effectiveness of the proposed course suggestion models, we applied assessment tests with a metric based on average similarity scores. This evaluation methodology compared the academic performance of a fixed student to the courses preferred by the BERT, word2vec, and TF-IDF models. Overall, the average similarity was computed using cosine similarity metrics, and the student's course profile was graphically depicted against the courses recommended by the proposed models.

Moreover, we expanded the same dataset to our second type of dataset, the Coursera course dataset, which includes all extra courses apart from the ongoing dataset. Taking advantage of the larger set of course tuples, the thesis aims to assess how successful the course suggestion models are at suggesting the right course in light of the more extensive courses offered. This evaluation helped to determine the generalizability and possibility of scaling up the models of course recommendations for students in different subjects and specialties.

3.8 Performance check with the Coursera dataset

The models were further tested and verified on a separate dataset collected from Coursera to generalize the recommendation system. The same preprocessing technique was used for this dataset, the same feature extraction and similar clustering methods were used, and several recommendations were created for several of the students, proving that the models established here can work on any dataset.

Using the performance check with the Coursera course dataset, we learned the impressions of the course suggestion models as they endorse courses from a broader array of available course choices. This evaluation provided useful insights into the stability and generalizability of the model in practice, where students could have diverse ways to plan for their future and allocate their time and resources to their academic interests.

In evaluating the average similarity scores and assessments using the data from Coursera courses, the assessment of the course suggestion models offered a broad understanding of their performance and applicability. These assessments helped provide significant information on the quality and appropriateness of the proposed courses and helped to improve further and develop necessary recommendations for the models.

Algorithm 1: Course Suggestion Algorithm using Context-Aware Clustering and Similarity Scoring

Input

- $student\_course\_results$ is a dictionary containing the courses and marks for each student from NPTEL.
- $Courses$ is a list containing all available course names.
- df is a data frame containing topic information after context-aware clustering of courses.
- Coursera is a list containing course names from the Coursera dataset.

1. **Data Preparation**:

- $final\_df \leftarrow$ $\text{process\_folders\_and\_files}(\text{folder}_{\text{base}_{\text{path}}}, \text{folder\_names})$
- $df \leftarrow$ $final\_df['\ Name', '\ Course\ Id', '\ CourseName', '\ result']$

2. **BERTopic Model Context-Aware Clustering of Courses**:

$$similarity(A, B) = \frac{A.B}{\|A\|\|B\|}$$

Where:A and B are vectors representing two courses (in this case, BERT $embeddings$).

- $Topic\_df \leftarrow Course\_clusters(Courses)$
- Normalize the student scores.
- The data were structured with student names, course names, and normalized scores.

SVD Model Training:

- Factorize the student-course interaction matrix R into U, Σ, V^T

$$R \approx U.\Sigma.V^T$$

- U; Student latent factors.
- Σ: Singular values matrix.
- V^T: Course latent factors.

Predicting ratings using:

$$\hat{r}_{ui} = U_{u}.\Sigma \cdot V_i^T$$

U_u and V^T arestudent u and i latent factor vectors, respectively.

Prediction accuracies concerning the RMSE and MAE on the test set

3. **Course Suggestion Using Similarity Scoring**:

- $best\_course \leftarrow$ $Suggest\_Course\_Bert(student\_courses, df)$

4. **TF-IDF with Agglomerative Clustering**:

$$TF - IDF(t, d) = TF(t, d) \times IDF(t)$$

Where:

$TF(t, d)TF(t, d) is\ the\ term\ frequency\ of\ term\ tt\ in\ document$

$dd.$

$IDF(t)IDF(t)\ is\ the\ inverse\ document\ frequency\ of\ term\ tt.$

- $X, vectorizer \leftarrow$ $vectorize\_courses\_tfidf(Courses)$
- $labels, clustering\_model \leftarrow$ $perform\_clustering\_tfidf(X)$
- $suggested\_course \leftarrow$ $suggest\_course\_for\_student\_tfidf(student_{course_{result}},$

$vectorizer, clustering\_model, labels,$

$Courses)$

5. **Word2vec with K-Means**:

$$J = \sum_{i=1}^{k} \sum_{x \in S_i} \|x - c_i\|$$

where$J\ is\ the\ objective\ function\ to\ minimize.$

$k\ is\ the\ number\ of\ clusters.$

$Si\ is\ the\ set\ of\ points\ assigned\ to\ cluster\ i.$

$ci\ is\ the\ centroid\ of\ cluster\ ii.$

$\|\cdot\|\ 2 denotes\ the\ Euclidean\ distance.$

- $model \leftarrow create\_word2vec\_model(Courses)$
- $vectors \leftarrow vectorize\_courses(Courses, model)$

6. **Performance Comparisons**:

- $similarity\_bert \leftarrow$ $calculate\_avg\_similarity(student_{course_{result}},$

$$Suggest\_Course\_Bert(student\_courses, df))$$

- $similarity\_word2vec \leftarrow calculate\_avg\_similarity(student_{course_{result}},$

$$predict\_word2vec(student\_course\_result, Courses))$$

- $similarity\_tfidf \leftarrow calculate\_avg\_similarity(student_{course_{result}},$

$$predict\_tfidf(student\_course\_result, Courses))$$

7. **Performance Plots**

- $plot\_student\_similarities(student_{course_{results[\text{Swathi}]}},$

$$df, Courses)$$

8. **Performance Check with the Coursera Courses Dataset**:

- $Coursera\_df \leftarrow load\_coursera\_dataset()$
- $Coursera \leftarrow Coursera\_df["CourseName"].tolist()$
- $df1 \leftarrow Course\_clusters(docs = Coursera)$
- $plot\_student\_similarities(results = student\_course\_results['KeerthanJamanjyothi'], df = df1, Courses = Coursera)$

4. Results

In this section, we discuss the findings of the performed research into course suggestion models and their assessment based on three methodologies. In this case, TF-IDF, Word2Vec, and BERT can be distinguished as preprocessing methods for textual data for classification purposes. This study aimed to establish the efficiency of these models in providing recommendations for courses close to the student's previous course history.

4.1 BERTopic-SVD Model Evaluation Metrics

This study evaluates SVD, and from the results, one can determine how SVD works to improve educational course recommendations given student interactions. Based on the results found in the present study, the SVD model achieved desirable performance in terms of student rating prediction while remaining significantly accurate. The choice of the evaluation parameters, such as the root mean square error (RMSE) and the mean absolute error (MAE), was equally effective and produced accurate rating discrepancies with only a slight variation from the actual interaction student course ratings.

Table 1. Values of RMSE and MAE for the test set and cross-validation

	RMSE	**MAE**
Test set	1.3267	1.0802
Cross-Validation	1.3518(average)	1.08700(average)

Table 2. Evaluating the RMSE and MAE of the SVD algorithm on 5 splits

	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Mean**	**Std**
RMSE (testset)	1.3693	1.3472	1.3188	1.3178	1.4059	1.3518	0.0332
MAE (testset)	1.0975	1.0802	1.0805	1.0721	1.1047	1.0870	0.0121
Fit time	0.18	0.19	0.19	0.19	0.19	0.19	0.01
Test time	0.02	0.03	0.03	0.02	0.02	0.02	0.00

Table I and Table II list the metrics for the BERTopic-SVD model. The metrics and evaluations presented in the study offer a clear picture of the effectiveness of the singular value decomposition (SVD) algorithm in the context of educational course recommendation systems. The table shows the test set's root mean square error (RMSE) and mean absolute error (MAE). An RMSE of 1.3267 and an MAE of 1.0802 indicate that, on average, the predicted ratings are off by approximately 1.33 and 1.8 points from the actual ratings, respectively. These values suggest a fairly accurate identification of students'course preferences and performance.

Cross-validation is performed across the five folds to provide more information on the algorithm's stability. The average RMSE of 1.3518 and MAE of 1.0870 across folds confirm that the algorithm is consistent across different partitions of the data and, therefore, has the potential to perform well on unseen data. The small standard deviations of 0.0332 for the RMSE and 0.0121 for the MAE show little variation in the prediction errors, proving that the SVD approach is accurate for recommending courses. Fig. 2 shows the values of the RMSE and MAE in graphical form.

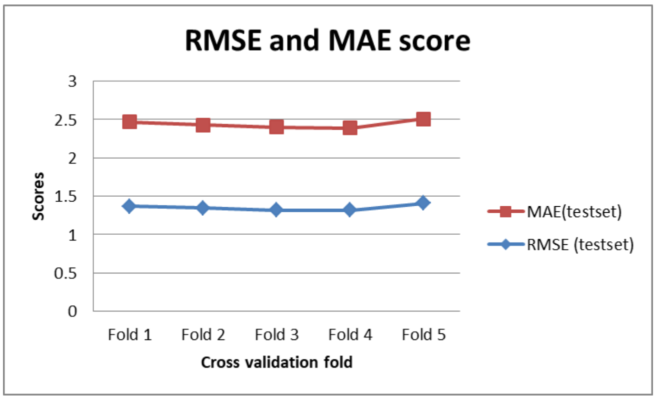

Fig 2: RMSE and MAE values for the test sets for different folds

4.2 Similarity score

To assess the model's effectiveness, we computed the cosine similarity between the student details of a particular course and the details suggested by the models concerning the Nptel students' dataset and another dataset, such as the Coursera dataset. The outcomes might help to identify the strengths and weaknesses of the paradigm of comparing models in terms of semantic correspondence and determine several courses of interest to various learners based on their enrollment history and preferences.

Table 3. Model performance table

Student	Nptel dataset			Coursera dataset		
	TFIDF model	**Word2Vec model**	**BERT model**	**TFIDF model**	**Word2Vec model**	**BERT model**
Student 1	0.83	0.86	0.89	0.86	0.88	0.88
Student 2	0.84	0.83	0.91	0.87	0.91	0.91
Student 3	0.82	0.82	0.86	0.80	0.83	0.86

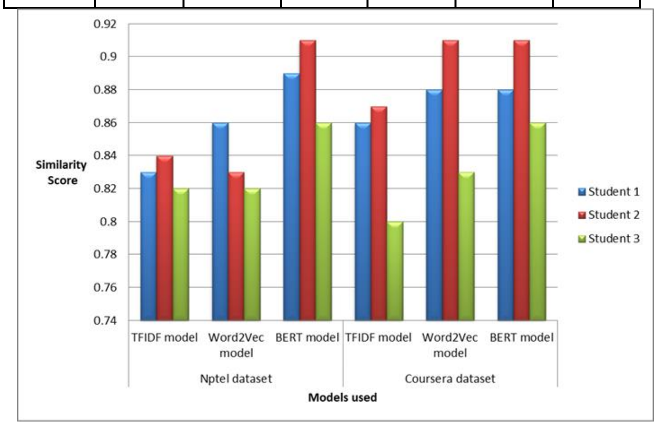

Fig 3 Performance graphs of the models for both the student detail dataset and the Coursera dataset

The table shows the average similarity scores between the students' course detail dataset and the suggested courses from three models. In text analysis, we identified three preprocessing techniques that achieved the highest average score: TF-IDF, word2vec, and BERT. The course detail dataset containing the details of the students was compared with the Coursera dataset to analyze the performance of the course suggestion models.

In the case of Student 1, the best result across the three models was obtained using BERT, with a mean similarity value of 0.89, meaning that the suggested courses were closest to the courses in their detailed dataset according to the tuned model compared to other models. Similarly, for Students 2 and 3, the BERT model was rated higher than both the TF-IDF and word2vec models, where it attained higher mean similarity scores.

In general, the BERT model performed slightly better in recommending courses that aligned with students' past course information for most of the students. This means utilizing pure semantic vectors and sophisticated language models such as BERT can capture the contextualized aspects of course titles and improve course suggestions. Nevertheless, it is crucial to note that the performance can be different in different cases due to the unique features of the given datasets and the level of difficulty of the course titles.

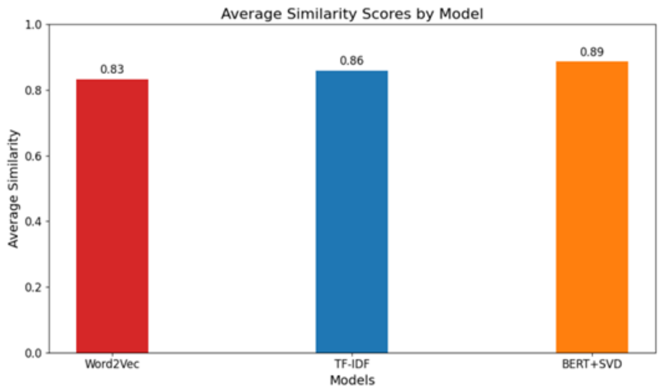

Fig 4: Student 1 performance plot on the Nptel dataset

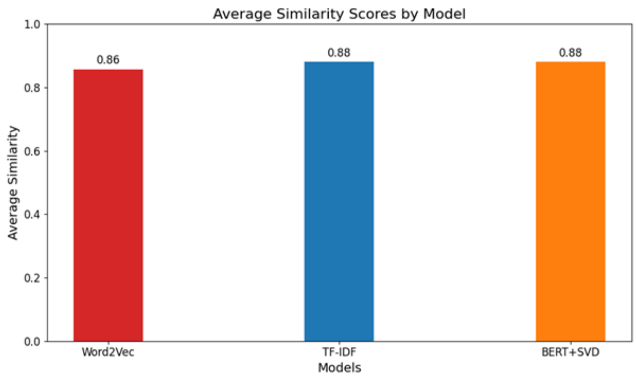

Fig 5 Student 1 performance plot on the Coursera dataset

Figs. 4 and 5 show the graphical representation of model performance for student 1 for the Npteland Coursera datasets, respectively. The BERT model performed the best among the three models.

By comparing the performance of these recommender systems, this study emphasizes the need to apply various techniques and methods to provide accurate recommendations for pertinent courses to students given their previous records. Moreover, validating and refining the integration of the models with the current data and future student-defined data may require additional adjustments.

5. Discussion

The findings of our work present valuable theoretical contributions to the domain of personalized course recommendation frameworks and the application of BERT SVD-based models. The evaluation results of the average similarity score further show how the BERT model outperforms the other models in offering recommendations for courses most similar to those the students previously took.

The results presented compare various recommendation techniques based on their performance metrics, which is why some of the most commonly used accuracy assessment indices include the root mean square error (RMSE) and mean absolute error (MAE). According to these techniques, BERTopic modeling integrated with singular value decomposition (SVD) yields the lowest RMSE of 1.3267 and an MAE of 1.087.

Table 4. Comparison of the model with existing work

Ref no	Technique	RMSE	MAE
[11]	k-means clustering algorithm + Collaborative filtering	1.8998089	1.133
[18]	KNN baseline	2.72	1.9601
	SVD	2.6161	1.9226
[21]	Constrained Matrix Factorization	4.5320 ± 0.0022	6.6764 ± 0.0029,
Our model	BERTopic modeling+SVD	1.3267	1.087

This means that the BERTopic-SVD model achieves better performance in terms of predicting the relevant course recommendations compared to the other approaches that were tested and include k-means clustering with collaborative filtering[11], the KNN baseline[18], SVD alone[18], and constrained matrix factorization[21].

From this table, it can be inferred that the BERTopic-SVD model has lower RMSE and MAE values than the other models, which gives better predictions of students' preferences by the courses they enrolled in and the performance they achieved. This approach uses BERT embeddings in course clustering through topic modeling and SVD to filter courses based on student preferences, which captures the fine-grained similarity between courses and students. Such outcomes prove the efficiency of combining complex NLP techniques with the collaborative filtering approach to increase the rate of course recommendation, which represents a student's profile.

This highlights the effectiveness of using advanced language models in analyzing semantic similarity and increasing the personalized course recommendations needed for efficient learning. However, it is critical to recognize the current body of work that stresses the importance of conducting a thorough comparative evaluation of different recommendation methods. This valuable research contributes to the existing knowledge base by detailing these models' strengths and limitations. It can eventually serve as guidelines for further exploration based on new improvements to personalized course recommendation systems. In this discussion, we emphasize the importance of BERT-based models for enhancing recommendation model performance and user satisfaction; however, we also recognize future work for more development in this emerging field.

Despite its advantages, this research has several drawbacks. The studies presented also have their drawbacks. First, matches were considered based only on course titles; thus, it can be assumed that other elements—such as course descriptions or prerequisites—were not considered. Moreover, the evaluation metrics are mainly linked to average similarity scores; however, using these metrics may not account for the differences in students' preferences or individual approaches to learning. Similarly, we have not incorporated a temporal aspect into the course popularity and may be insensitive to changes in curriculum over the academic year and across semesters. To overcome these limitations, future research could involve other attributes and enhance the assessment techniques to prepare a new comprehensive matrix to grade the effectiveness of course recommendations.

6. Conclusion and Future Scope

This study assessed the effectiveness of three distinct course recommendation models—TF-IDF, word2vec, and BERT-SVD—employed and trained using historical course enrollment data from the Nptel. The proposed BERT SVD course recommendation system uses the similarity score and predicted rating via SVD to enhance the system output. The RMSE obtained for the proposed system is 1.3267, and the MAE is 1.080. After analyzing and comparing the TF-IDF and word2vec models, the value analysis revealed that the BERT model had higher similarity score scores than the other models concerning different student datasets in general and on average. This implies that BERT is better equipped to identify semantic relations more effectively and offers relevant course recommendations based on student preferences. These findings present evidence of the potential of highly developed language portraits such as BERT in enhancing educational technology domains and promoting customized learning processes. Through the ability to better comprehend the context and competition of courses based on their titles, BERT can facilitate more precise recommendations, the results of which will improve academic performance. Furthermore, our study is relevant because it highlights how various approaches to personalization and recommenders might work. The evaluation process should not be left out to likely clueless automatic testing but must be carried out meticulously to determine the best strategy in certain scenarios.

Therefore, there are some areas for future work and improvement to models for course recommendation. One possible area for further research is the opportunity to utilize the best features of several models, such as BERT, word2vec, and TF-IDF, to enhance the effectiveness and reliability of recommendations. However, incorporating user engagement/feedback and learning management systems that employ adaptive learning principles can also help improve and update course recommendation services based on users' dynamic needs, preferences, and academic performance. It is necessary to investigate the applicability of course recommendation models across different educational contexts and utilization spheres. Evaluating these models on large-scale and diverse datasets may provide useful insights into the effectiveness of the models. Furthermore, it becomes necessary to explore the ability to interpret the model's outcomes and the measures to avoid bias in the recommendations made on the course of the student's education. In retrospect, pursuing continued research and development of techniques that favor course recommendation serves as the possibility to decentralize education.

Conflict-of-Interest Statement

The authors indicate no conflicts of interest in publishing this work. They also have no financial assistance or interests that might create a conflict of interest for the work addressed in this paper.

References

[1] M. C. Urdaneta-Ponte, A. Mendez-Zorrilla, and I. Oleagordia-Ruiz, "Recommendation Systems for Education: Systematic Review," Electronics, vol. 10, no. 14, p. 1611, Jul. 2021. [Online]. Available: https://doi.org/10.3390/electronics10141611

[2] H. Ko, S. Lee, Y. Park, and A. Choi, "A Survey of Recommendation Systems: Recommendation Models, Techniques, and Application Fields," *Electronics*, vol. 11, no. 1, p. 141, Jan. 2022. [Online]. Available: https://doi.org/10.3390/electronics11010141

[3] B. Li *et al.*, "A personalized recommendation framework based on MOOC system integrating deep learning and big data," *Computers & Electrical Engineering*, vol. 106, p. 108571, Mar. 2023. [Online]. Available: https://doi.org/10.1016/j.compeleceng.2022.108571

[4] S. Amin, M. I. Uddin, A. A. Alarood, W. K. Mashwani, A. Alzahrani, and A. O. Alzahrani, "Smart E-Learning Framework for Personalized Adaptive Learning and Sequential Path Recommendations Using Reinforcement Learning," *IEEE Access*, vol. 11, pp. 89769–89790, Jan. 2023. [Online]. Available: https://doi.org/10.1109/access.2023.3305584

[5] G. Muzdybayeva, D. Khashimova, A. Amirzhanov, and S. Kadyrov, "A Matrix Factorization-based Collaborative Filtering Framework for Course Recommendations in Higher Education," Jun. 2023. [Online]. Available: https://doi.org/10.1109/icecco58239.2023.10147152

[6] K. K. Jena *et al.*, "E-Learning Course Recommender System Using Collaborative Filtering Models," *Electronics*, vol. 12, no. 1, p. 157, Dec. 2022. [Online]. Available: https://doi.org/10.3390/electronics12010157

[7] M. Premalatha, V. Viswanathan, and L. Čepová, "Application of Semantic Analysis and LSTM-GRU in Developing a Personalized Course Recommendation System," *Applied Sciences*, vol. 12, no. 21, p. 10792, Oct. 2022. [Online]. Available: https://doi.org/10.3390/app122110792

[8] G. M. Dhananjaya, R. H. Goudar, A. Kulkarni, V. N. Rathod, and G. S. Hukkeri, "A Digital Recommendation System for Personalized Learning to Enhance Online Education: A Review," *IEEE Access*, p. 1, Jan. 2024. [Online]. Available: https://doi.org/10.1109/access.2024.3369901

[9] J. Wang, H. Xie, O. T. S. Au, D. Zou, and F. L. Wang, "Attention-Based CNN for Personalized Course Recommendations for MOOC Learners," Aug. 2020. [Online]. Available: https://doi.org/10.1109/iset49818.2020.00047

[10] J. Li and Z. Ye, "Course Recommendations in Online Education Based on Collaborative Filtering

Recommendation Algorithm," *Complexity*, vol. 2020, pp. 1–10, Dec. 2020. [Online]. Available: https://doi.org/10.1155/2020/6619249
[11] B. Mondal, O. Patra, S. Mishra, and P. Patra, "A course recommendation system based on grades," *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Mar. 2020. [Online]. Available: https://doi.org/10.1109/iccsea49143.2020.9132845
[12] Y. Lin, S. Feng, F. Lin, W. Zeng, Y. Liu, and P. Wu, "Adaptive course recommendation in MOOCs," *Knowledge-based Systems*, vol. 224, p. 107085, Jul. 2021. [Online]. Available: https://doi.org/10.1016/j.knosys.2021.107085
[13] T. Morrow, A. R. Hurson, and S. S. Sarvestani, "Algorithmic Support for Personalized Course Selection and Scheduling," Jul. 2020. [Online]. Available: https://doi.org/10.1109/compsac48688.2020.00027
[14] D. B. Guruge, R. Kadel, and S. J. Halder, "The State of the Art in Methodologies of Course Recommender Systems—A Review of Recent Research," *Data*, vol. 6, no. 2, p. 18, Feb. 2021. [Online]. Available: https://doi.org/10.3390/data6020018
[15] Q. Li and J. Kim, "A Deep Learning-Based Course Recommender System for Sustainable Development in Education," *Applied Sciences*, vol. 11, no. 19, p. 8993, Sep. 2021. [Online]. Available: https://doi.org/10.3390/app11198993
[16] V. A. Nguyen, H.-H. Nguyen, D.-L. Nguyen, and M.-D. Le, "A course recommendation model for students based on learning outcome," *Education and Information Technologies*, vol. 26, no. 5, pp. 5389–5415, Apr. 2021. [Online]. Available: https://doi.org/10.1007/s10639-021-10524-0
[17] H. Wang and W. Fu, "Personalized Learning Resource Recommendation Method Based on Dynamic Collaborative Filtering," *Journal on Special Topics in Mobile Networks and Applications/Mobile Networks and Applications*, vol. 26, no. 1, pp. 473–487, Oct. 2020. [Online]. Available: https://doi.org/10.1007/s11036-020-01673-6
[18] W.-E. Kong, S.-C. Haw, N. Palanichamy, and S. H. A. Rahman, "An e-Learning Recommendation System Framework," *International Journal on Advanced Science, Engineering and Information Technology/International Journal of Advanced Science, Engineering and Information Technology*, vol. 14, no. 1, pp. 10–19, Feb. 2024. [Online]. Available: https://doi.org/10.18517/ijaseit.14.1.19043
[19] M. Liang and T. Niu, "Research on Text Classification Techniques Based on Improved TF-IDF Algorithm and LSTM Inputs," *Procedia Computer Science*, vol. 208, pp. 460–470, Jan. 2022. [Online]. Available: https://doi.org/10.1016/j.procs.2022.10.064
[20] R. Wang and Y. Shi, "Research on application of article recommendation algorithm based on Word2Vec and Tfidf," *2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA)*, Feb. 2022. [Online]. Available: https://doi.org/10.1109/eebda53927.2022.9744824
[21] S.-T. Zhong, L. Huang, C.-D. Wang, and J.-H. Lai, "Constrained Matrix Factorization for Course Score Prediction," Nov. 2019. [Online]. Available: https://doi.org/10.1109/icdm.2019.00199