International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://theaspd.com/index.php

A Data-Driven Framework Based on Machine Learning Approaches for Restructuring Computer Science Courses

Ritika Awasthi,Dr C.V. Raman University Bilaspur , Chhattisgarh, India
(ritikaawasthi1612@gmail.com)
Dr. Abhinav Shukla, Dr. C. V. Raman University, Bilaspur, Chhattisgarh, India
(amitshukla9611@gmail.com)

Abstract: *This study proposes an approach to machine learning for changing the computer science curriculum at higher education institutions. The approach combines topic modeling, sentiment analysis, clustering, regression, and algorithmic recommendation to derive actionable insights for curriculum design by examining course content, student performance, feedback, and institutional data. The research indicates that this method enhances curriculum relevance, customizes student learning directions, and corresponds with Program Specific Outcomes (PSOs) and Program Educational Objectives (PEOs). Simulation performed using a prototype LMS dashboard confirms the model's viability and applicability. The findings demonstrate how a data-driven framework can proficiently connect educational resources with learner expectations while accomplishing the institutional goals.*
Keywords: *Curriculum Design, Machine Learning, Topic Modeling, Sentiment Analysis, Student Segmentation, Recommender Systems, Program Specific Output (PSO), Program Educational Objectives (PEO), Learning Management System (LMS)*

INTRODUCTION

The swift development of the technology sector exerts mounting pressure on academic institutions to guarantee that their computer science curricula are pertinent, thorough, and in accordance with industry benchmarks and student requirements. As novel programming paradigms, development tools, and interdisciplinary applications rapidly emerge, the traditional curriculum frequently comes short, neglecting to integrate the most recent technologies and methodologies. This misalignment generates graduates who may lack the essential, job-ready competencies anticipated by employers, hence exacerbating the disparity between academic training and professional expectations(Sekiya et al., 2015). Furthermore, traditional curriculum creation techniques often adhere to inflexible frameworks, with few updates generally influenced by expert agreement rather than empirical data. These techniques frequently neglect individual learner variances, student feedback, and changing performance patterns. As a result, numerous computer science programs experience obsolete course material, redundancy among modules, and inadequate personalization, resulting in student disengagement, inconsistent learning outcomes, and underutilization of educational resources. This study provides a comprehensive, data-driven methodology utilizing machine learning and educational data mining to automate and enhance curriculum creation. The suggested approach analyzes many elements of educational data, such as course syllabi, student performance indicators, feedback sentiment, and institutional constraints, to dynamically adjust curricular structures in alignment with current academic objectives and future industrial demands. The above approach provides ongoing improvement of the curriculum, supports individualized learning trajectories, enhances student happiness, and aligns educational results with established Program Specific Outcomes (PSOs) and Program Educational Objectives (PEOs)(Jha, 2023).Mapping Curriculum Insights to PSOs and PEOs Within the scheme of things that has been proposed, the machine learning components are strategically aligned with the Program Specific Outcomes (PSOs) and Program Educational Objectives (PEOs) of the institution. PEO1 (Strong foundation in computer and analytical skills) and PSO1 (Domain Knowledge) are both supported by topic modeling, which identifies significant academic subjects and guarantees that fundamental content is covered. Through the analysis of student suggestions and the provision of information concerning content improvements, sentiment analysis contributes to both PEO3 (Responsiveness to student needs and quality assurance) and PSO2 (Curriculum relevance and innovation). By segmenting students in order to provide them with individualized support and learning directions, algorithms for clustering provide support for PEO4,

which stands for personalized and student-centered education. Predictive regression models support the implementation of PSO2 and PEO2 (Effective academic advice and student support) by predicting trends in student performance and making it achievable to carry out early interventions. The last point is that recommender systems encourage lifelong, adaptive learning that is in line with PSO4 (Self-directed and continuous learning), as well as PEO2 and PEO4, by recommending classes that are tailored to an individual's specific interests and long-term professional objectives. Through the use of this mapping, it is demonstrated that the data-driven approach not only improves the curriculum in terms of its structure, but also contributes to the achievement of bigger educational goals and outcomes.

Methodology: This work utilizes a systematic machine learning-based approach for improving curriculum planning in computer science education. The approach utilizes both organized and unstructured data from academic institutions for assessing and rewriting course offerings, anticipating student performance, and recommending personalized academic pathways.

Data Source	Description
Student Performance Data	Includes grades, attendance rates, and assignment submissions. Reflects student engagement and academic success.
Course Syllabi	Collected from five different institutions to ensure diversity in content, structure, and thematic focus.
Student Feedback	Gathered from course evaluations and academic platforms. Covers sentiment on content, teaching quality, and practical relevance.
Institutional Data	Includes course availability, faculty profiles, class sizes, and departmental learning objectives.

Table 1: Data Sources Table

Preprocessing strategies:

- Various strategies were used to prepare the dataset for machine learning models. Attendance and grade data from 50 students were normalized using Min-Max Scaling. Scaling attendance rates from 62% to 98% to 0–1 ensured homogeneous input for regression and grouping methods.
- More than 10 course syllabi and 200 student feedback items were vectorized using TF-IDF for topic modeling and Word2Vec for document similarity analysis. Key instructional topics and sentiment context were collected.
- One-hot encoding was used for student profiles (UG/PG level, elective selections) and course types (Core, Elective, Lab).

Table 2.1: Sample of Preprocessed Student Dataset

Student ID	Grade (%)	Attendance (%)	Scaled Grade	Scaled Attendance	UG/PG	Course Type
S101	82	91	0.78	0.89	UG	Core
S102	67	75	0.56	0.60	UG	Elective
S103	92	98	0.92	1.00	UG	Core
S104	74	62	0.65	0.42	UG	Lab

Implemented Machine Learning Techniques:

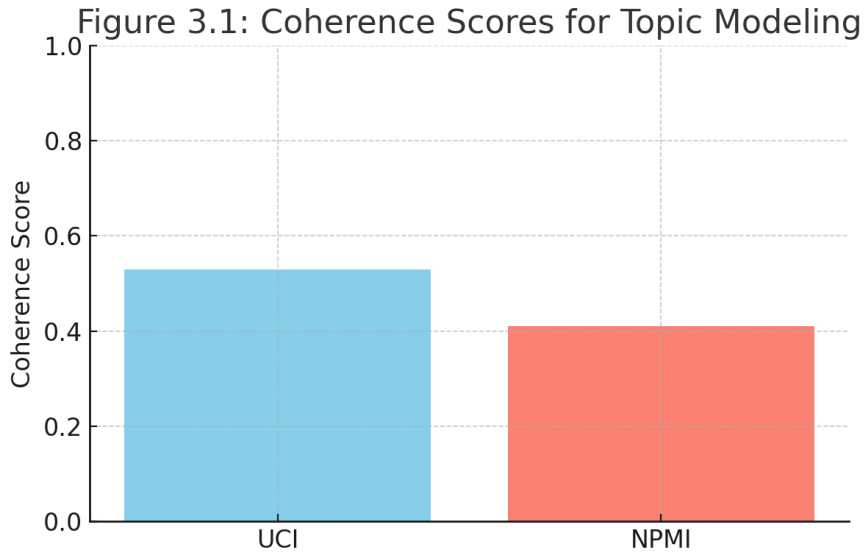
Topic Modeling (LDA, NMF): Topic modeling methodologies were employed to categorize course syllabi into significant topics, including Artificial Intelligence, Data Science, and Web Technologies. This facilitated the identification of thematic deficiencies and redundancies within the program. Courses such as “Machine Learning” and “Neural Networks” consistently emerged within the theme cluster designated ‘AI and Data-Driven Systems,’ signifying a coherent subject matter.

Sentiment analysis was performed on student comments with Support Vector Machine (SVM) and Naïve Bayes classifiers(Tao et al., 2023). The SVM attained an F1-score of 0.84, markedly surpassing Naïve Bayes. The investigation indicated adverse opinion toward antiquated theory-based modules and underscored favorable sentiment towards practical lab sessions and contemporary content, which directly impacted course changes.Clustering (K-Means, Hierarchical): Clustering techniques were employed to categorize pupils according to academic indicators, including grades, attendance, and assignment completion. PCA representations of the clustering outcomes identified three separate categories of students: high-performing, average, and at-risk. These clusters were utilized to provide specialized academic courses and focused support initiatives.Regression Models (Linear Regression, Random Forest): Regression analysis was utilized to forecast student performance based on prior academic data. The Random Forest model surpassed Linear Regression, with a R^2 value of 0.81. The model determined that attendance and punctual assignment submission were the primary determinants of final course grades.Recommender Systems (SVD, Cosine & Jaccard Similarity): Collaborative and content-based recommender systems were created to propose pertinent electives depending on student performance and interests. The algorithm attained an accuracy of over 85%, frequently recommending "Deep Learning" to students who completed "Machine Learning," thereby facilitating controlled educational progression and curriculum consistency.

RESULT AND ANALYSIS

3.1 Topic Modeling

Course syllabi were evaluated using LDA and NMF to identify established themes, including AI, Web Development, Data Science, and Cybersecurity. These approaches helped educators find topic clusters by grouping keywords and course content. UCI (0.53) and NPMI (0.41) coherence scores revealed that the selected themes were semantically valid and logically coherent.

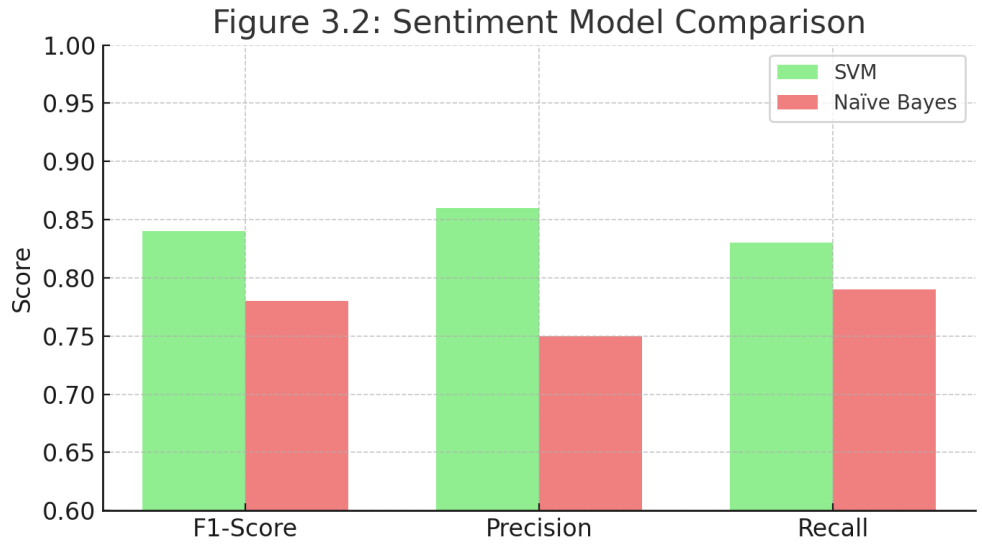


This research also identified redundant areas (e.g., numerous web programming basics courses) and underrepresented topics (e.g., ethical computing or cloud security), facilitating targeted curriculum redesign.

3.2 Sentiment Analysis

Two classification algorithms, SVM and Naïve Bayes, were employed to examine student comments(OSMANOĞLU et al., 2020). These models assessed feedback as favorable or negative. SVM surpassed Naïve Bayes with a higher F1-score of 0.84, demonstrating more accurate and balanced categorization.

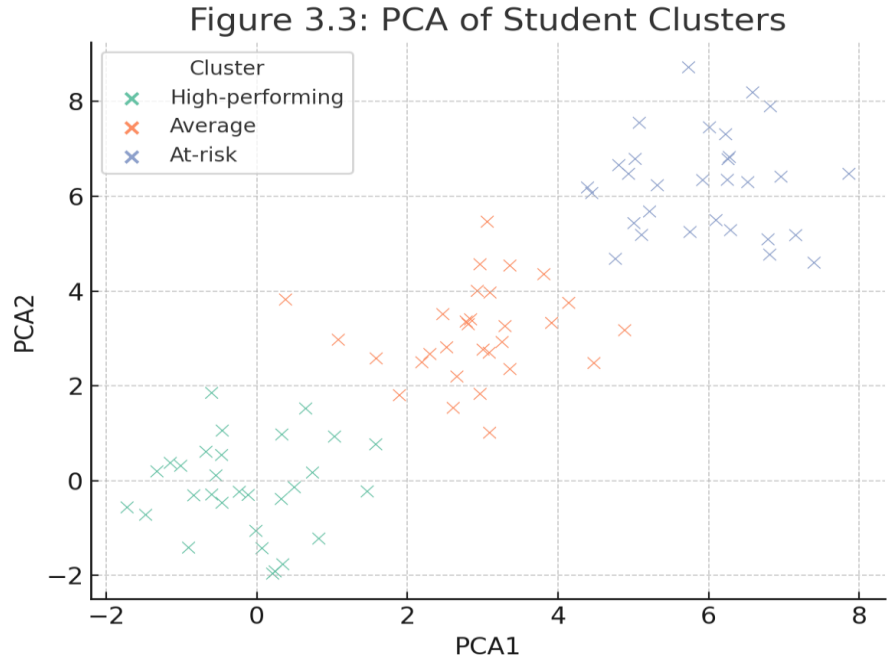
International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8s, 2025
https://theaspd.com/index.php



Positive opinion was shown for hands-on lab activities, interactive teaching methods, and updated tools, while negative sentiment was largely focused at obsolete theory modules and lack of real-world applicability. These insights serve educators evaluate what to make changes in and which material to preserve.

3.3 Student Clustering

K-Means and Hierarchical Clustering were used to group students by academic performance criteria (grades, attendance, and assignment submissions) to identify learning patterns(Kausar et al., 2018). Results showed three clusters: high-performing, average, and at-risk students.

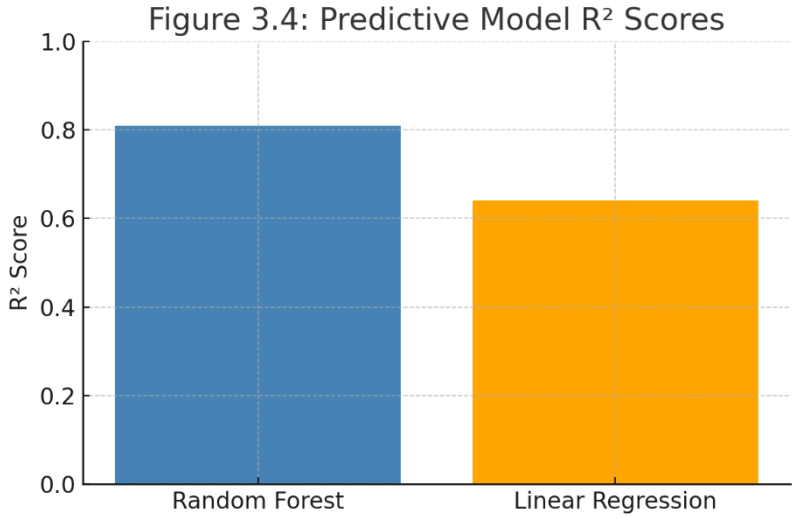


The clusters were represented in two dimensions using Principal Component Analysis (PCA), making performance variance easy to understand. The insights allow institutions to offer tailored academic support, such as advanced electives for top achievers and skill-bridging or remedial courses for underperformers.

3.4 Predictive Performance

Early academic indications were utilized to predict final grades using regression models. The Random Forest model outperforms Linear Regression, with a high R^2 score of 0.81 rather than 0.64. Random

Forest is capable of forecasting academic outcomes because it can explain 81% of student grade variance. Prior performance, attendance, and assignment completion influenced forecasts.



These predictions allow institutions to implement early interventions, alerting instructors to the students who may need more help.

3.5 Recommender Systems

A hybrid recommendation system was created combining Collaborative Filtering (SVD) and Content-Based Filtering (Cosine and Jaccard Similarity)(Baseera & Srinath, 2014). The collaborative filtering approach predicted student electives with 85% accuracy using previously enrolled patterns. To ensure logical course progressions, content-based filtering was utilized to assess course descriptions and recommend "Deep Learning" to students who completed "Machine Learning".

Course Completed	Recommended Course	Match Accuracy (%)
Machine Learning	Deep Learning	89
Web Dev	Web Security	82
Data Structures	Algorithms	86

Table 2: Course Recommendation and Match Accuracy

This minimized content redundancy and created structured, customized learning paths for each student's skill development.

3.6 LMS Dashboard Simulation

A prototype dashboard has been developed to illustrate practical use by mimicking integration with LMSs like Moodle and Blackboard(Marmoah et al., 2023). The dashboard showed real-time academic indicators like individualized course recommendations, skill gaps, and updated course suggestions. If a student experienced trouble with programming assignments, the dashboard would advise an introductory Python module or extra practice.

Feature	Description
Recommended Courses	Suggests the next courses based on performance
Skill Gap Alerts	Identifies areas where students need help
Progress Tracker	Monitors learning progress over time
Remedial Resources	Provides extra learning material for weak topics

Table 3: LMS Dashboard Features and Description

This adaptive interface provides immediate, information-driven guidance to optimize learning and enhance academic success.

Mapping with PSO and PEO

Within the framework that has been proposed, the machine learning components have been carefully coordinated with the Program Specific Outcomes (PSOs) and Program Educational Objectives (PEOs) of

the institution. With the identification of essential academic areas and the guarantee of basic curriculum coverage, topic modeling assists PSO1 and PEO1(Aftabuzzaman & Wahr, 2021). The process of sentiment analysis provides an improvement to PSO2 and PEO3 by capturing the perspectives of students and improving the responsiveness of the curriculum. In alignment with PEO4, clustering algorithms make it possible to provide individualized academic help. Early interventions that are in keeping with PEO2 have been made possible by the use of predictive regression models, which anticipate structure in student performance. In the final analysis, recommender systems facilitate adaptive learning as well as structured academic growth, so satisfying not only PEO2 and PEO4 but also PSO4 requirements.

Model/Insight	**PSO**	**PEO**	**Explanation**
Topic Modeling	PSO1	PEO1	Identifies curriculum domains and redundancies, ensuring core competency alignment.
Sentiment Analysis	PSO2	PEO3	Analyzes feedback to refine curriculum relevance and responsiveness.
Clustering & Segmentation	PSO3	PEO4	Segment learners for personalized learning support and academic scaffolding.
Predictive Models (Regression)	PSO2	PEO2	Forecasts academic risk and informs early intervention strategies.
Recommendation Systems	PSO4	PEO2, PEO4	Suggests courses aligned with skills and career goals for adaptive learning.

Table 3: Mapping with PSO and PEO

According to this mapping, it is clear that the data-driven approach not only improves the design of the curriculum but also benefits to the achievement of broader educational objectives.

CONCLUSION AND FRAMEWORK

The conclusion and recommendations for future research are as follows: this study demonstrates that machine learning has the potential to effectively enhance curriculum by aligning course content with learner profiles and the objectives of the institution. Integration of the system with actual learning management systems (LMS) platforms, expansion of the dataset across different institutions, incorporation of advanced natural language processing models, and evaluation of the system's long-term academic influence through longitudinal studies are all examples of future work.

REFERENCES

1. Aftabuzzaman, M., & Wahr, F. (2021). A comparative analysis of student learning experience in face-to-face vs. fully-online. *9th Research in Engineering Education Symposium and 32nd Australasian Association for Engineering Education Conference, REES AAEE 2021: Engineering Education Research Capability Development*, *1*. https://doi.org/10.52202/066488-0059
2. Baseera, & Srinath. (2014). Design and development of a recommender system for E-learning modules. *Journal of Computer Science*, *10*(5). https://doi.org/10.3844/jcssp.2014.720.722
3. Jha, P. (2023). Significance of Bloom's Taxonomy for Attainment of Program Outcome (PO) and Course Outcome (CO) in Educational Institute. *Journal Homepage: Https://Ejournal. Jhamobi. Com*, *August*.
4. Kausar, S., Huahu, X., Hussain, I., Wenhao, Z., & Zahid, M. (2018). Integration of Data Mining Clustering Approach in the Personalized E-Learning System. *IEEE Access*, *6*, 72724–72734. https://doi.org/10.1109/ACCESS.2018.2882240
5. Marmoah, S., Sukmawati, F., Poerwanti, J. I. S., Supianto, Yantoro, & Duca, D. S. (2023). Teacher Challenges in Designing the Learning after Curriculum Change: An Analysis of Learning Management System. *International Journal on Advanced Science, Engineering and Information Technology*, *13*(6). https://doi.org/10.18517/ijaseit.13.6.19655

6. OSMANOĞLU, U. Ö., ATAK, O. N., ÇAĞLAR, K., KAYHAN, H., & CAN, T. (2020). Sentiment Analysis for Distance Education Course Materials: A Machine Learning Approach. *Journal of Educational Technology and Online Learning*, *3*(1), 31–48. https://doi.org/10.31681/jetol.663733
7. Sekiya, T., Matsuda, Y., & Yamaguchi, K. (2015). Curriculum analysis of CS departments based on CS2013 by simplified, supervised LDA. *ACM International Conference Proceeding Series*, *16-20-March-2015*, 330–339. https://doi.org/10.1145/2723576.2723594
8. Tao, X., Shannon-Honson, A., Delaney, P., Dann, C., Xie, H., Li, Y., & O'Neill, S. (2023). Towards an understanding of the engagement and emotional behaviour of MOOC students using sentiment and semantic features. *Computers and Education: Artificial Intelligence*, *4*. https://doi.org/10.1016/j.caeai.2022.100116