Predicting Student Academic Performance using Feature Engineering on E-Learning Platforms

Utkarsh Loyalka
School of CS&IT
Jain (Deemed-to-be-University)
Bangalore, India
loyalkautkarsh@gmail.com

Sadeel Shamim Wani
School of CS&IT
Jain (Deemed-to-be-University)
Bangalore, India
Sadeelwani07@gmail.com

Sarthak Singh
School of CS&IT
Jain (Deemed-to-be-University)
Bangalore, India
sarthak3283@gmail.com

Dr. Suma S
School of CS&IT
Jain (Deemed-to-be-University)
Bangalore, India
suma@jainuniversity.ac.in

***Abstract*— The widespread adoption of e-learning platform has transformed modern education by enabling continuous monitoring of students engagement, learning behavior, and academic performance. Learning management system [LMS]. Such as moodle, Coursera and Edx collect large volume of behavioral data including login, activity, resource interaction, assignment, submission and discussion. Forum Participation. These datasets provide valuable insights that can be analyzed using machine learning algorithms to predict student academic outcomes and identify learners at risk of academic failure. However, raw LMS interaction data is often noisy, inconsistent and difficult to interrupt, which limits the Reliability of predictive models Feature Engineering plays a critical role in transforming raw behavioral logs into meaningful indicators such as study consistency, time on tasks, participation intensity, and learning persistence Students using real world data. Let's demonstrate that Engineered features. Significantly improved predicting, accuracy and interpretability of machine learning Model. This research analyzes how feature engineering enhances academic performance predicting models while maintaining transparency. Fairness and ethical AI adoption in education. The study synthesizes binding from recent research to propose a conceptual framework that supports Interpretable predictive analytics aligned with responsible AI principles in educational environment.**

***Keywords—component, formatting, style, styling, insert** (key words)*

I. Introduction

The Integration of digital technologies in education has led to the rapid expansion of online learning platform and learning management system elements. These systems record detailed information about student interaction with course materials, assignments, and. Collaborative activities As a result, education institutions now have access to Extensive datasets that capture students behavioral patterns throughout the learning process.

Education, Data Mining, EDM and learning Analytics has emerged as important research area that aim to extract knowledge from these digital traces. Improve teaching and learning outcomes. Machine learning algorithms are increasingly Used to analyze LMS interaction log in the order to predict student academic performance, Identify at risk learners and productive early. Interventions, strategies.

However, elements stares at typically consists of large volumes of raw clickstream data that include low level interaction events such as page view. Uh, file downloads and navigation logs. These raw logs are often unstructured and contained. redundant or irrelevant events that do not directly represent meaningful learning behavior. Without appropriate preprocessing, predictive models trained on such data may produce unreliable results.

Feature Engineering addresses test challenge by transforming raw elements interaction data into meaningful variables that represent student engagement, Study Habits. And performance patterns. Example of such features include logging frequency, time spent on learning materials, assignment submission behavior, and discussion forum participation. Research has shown that machine learning models trained on engineered feature outperform models trained on More data in both predictive accuracy and interpretability.

By Integrating feature engineering with machine learning techniques, educational institutes can develop predicting analytics system that provides actionable insights for instructors and support data-driven decision making in academic environments.

II. Statement of Problem

Despite the Availability of large volumes of a student interaction data from e-learning platforms developing reliable academic performance predictive models remain the major challenge. Raw elements datasets often contain noisy interaction locks, missing values, and inconsistent behavioral patterns that make it difficult to extract meaningful insights directly.

Another challenge arises from the complexity Of machine learning models used in predictive analytics. Advance has algorithms such as assemble models and deep learning architectures often operate as black boxes, making it difficult for educators to understand how predictions are generated. This lack of transparency reduces trust in predictive systems and limits their their adoption in educational institutes.

Furthermore, predicting models are unintentionally introduce Bias if the engineered features reflect demographic inequalities or unequal access to digital resources. Ethical concerns such as fairness, accountability and transparency must therefore be considered. And designing predictive analytics systems for education.

Feature engineering offers a potential solution by transforming raw LMS interaction logs into interpretable behavior indicators that reflect actual learning patterns. Properly engineered features improve prediction accuracy while also enabling Explainable machine learning models that educators can understand and trust.

III. OBJECTIVE OF THE STUDY

The main objective of this research is to examine how feature engineering contribute to improving the accuracy and interpretability of student academic performance. Prediction Models in eLearning environments.

The specific objectives of this study are:

1. To analyze the role of future engineering in improving machine learning prediction accuracy for student performance.

2. To identify key behavioral indicators Derived from LMS datasets that influence academic outcomes.

3.To examine machine learning techniques used for predicting student performance using engineer features.

4.To explore how explainable AI techniques improve transparency and trust in predictive models.

5.To propose a conceptual framework that integrates feature engineering and interpretable machine learning for responsible educational analytics.

IV. LITERATURE REVIEW

[1] Boujmiraz Darhmaoui and Drissi Conducted A comprehensive review of machine learning. And explainable AI techniques used for predicting student performance in educational environments. This study analyzed multiple research works that applied machine learning algorithms to. Elements datasets and emphasized the importance of reaching engineering and transforming law student interaction logs into meaningful behavioral indicators. The research highlights that engineered features such as engagement patterns, history. Consistency and learning persistence It significantly improves the accuracy and interoperability of predicting models.

[2] Angeioplastis et al. Proposed data-driven framework for predicting student academic performance using educational data mining techniques. The study analyzed real-world LMS datasets containing at Chadwick records of more than 450 university. Students across 9 semesters. Machine learning algorithms were applied to identify relationships between student behavior and academic outcomes. The finding demonstrate that engineered behavior features such as sports activity participation and assignment Completion pattern significantly improved predicting accuracy.

[3] Xu Explore the role of learning analytics and model linking student engagement in online learning environments. The study analyzed student interaction data collection from digital learning platforms and used machine learning techniques to identify engagement patterns to influence academic performances. The results indicates that students who demonstrate consistent engagement with learning materials are more likely to achieve higher academic success.

[4] Airlangga investigated the use of deep learning models for predicting The student academic Performances and educational datasets. Does this search evaluated multiple neural network architecture including convolution neural network, CNN and a long term memory Networks LSTM their education. The study found that combining features engineering were deep learning models significantly improve predicting accuracy and enables better understanding of each students learning behavior.

[5] Sandeepa, and Mahottala Evaluated various machine learning models. Predicting student academic performance using behavioral and. Demographic datasets. Their research compared multiple algorithms, including random forest support vector machines and logistic regression.

[6] A study published in the International Journal of Emerging Technology in Learning analyzed moodle LMS interaction logs to predict student academic performance using machine learning techniques. The research extracted behavioral features from elements logs, including log. Even frequency resources using patterns and assignment submission timings. The results showed that feature engineering significantly improves prediction accuracy compared to the traditional statistical methods.

RESEARCH METHODOLOGY

The study adopts a **qualitative research methodology supported by evidence from real-world datasets and recent academic research in educational data Minning and learning analytics.** Instead of conducting a new experimental dataset of conducting a new experimental dataset collection, the research Synthesizers finding from multiple peer reviewed studies that have implemented machine learning techniques for predicting student academic performance using learning management system [LMS] data.

The methodology focused primarily on understanding of **feature engineering transform raw behavior logs from e-learning platform into structured variables suitable for predicting models.** Many modern LMS platforms including Moodle, Blackboard and Canva. Record detailed logs of students interaction such as logging timestamps, pageviews, queues, tempts, forum post and assignment submission. These interaction logs generate large volume of behavioral data that can be analyzed using machine learning.

Several of these stories reviewed in the search relay on **real-world education datasets.** for example, Open University. The Open University learning analytics data sets (OULAD) Contains interaction data from **over 32,000 students and more than 10 million recorded LMS events**, making it one of the most comprehensive data sets used for educational data mining research. Researchers have used this data set to engineer behavioral features such as learning persistence, resource access pattern, and engagement frequency in order to predict student performance and course completion rates.

Similarly Angeioplastis et al. [2] analyzed 450 universities students across nine semesters using LMS behavioral records to develop predictive models capable of category categorizing students into performance groups such as high performing, average performing, and at risk learners.

The methodology used in this study involves the following steps:

1. **Literature identification-** Academic paper related to student performance Predictions, learning analytics and educational data mining were identified through academic database such as Science Direct, MDPI and arxiv.
2. **Dataset Examination-** The reviewed studies were analyzed to identify the type of data sets used, including LMS, interaction logs, academic grade records, and engagement metrics.
3. **Feature Engineering Analysis-** Each study was examined to determine how raw behavioral data was transformed into predictive features such as Skymont task login, frequency assignment, submission pattern, and participation. Collaboration learning activities. Model evaluation The machine learning algorithms used in such study, including random forests, support vector machine, SVM, logistic regression, and deep learning models work compared to evaluate their predictive performance.
4. Mosel Evaluation- The machine learning algorithm used in each study, including Random Forest, Support vector Machine (SVM) , Logisitic Regression and Deep learning models, were compared to evaluate their predictive performance.
5. Interpretability Assessment – The research also evaluated how interpretability Techniques such as SHAP values, Feature Importance analysis, and explainable AI frameworks help educators understand predictive results.

Through this structured methodology, the study provides a comprehensive understanding of how feature engineering improves both the accuracy and interpretability of academic prediction systems.

RESEARCH DESIGN

The Research design adopted in this study provides a comprehensive understanding of how feature engineering improves both the accuracy and interpretability of academic prediction systems.

The study begins by examining real-world implementations of predictive analytics systems used in higher education institutions. These systems utilize machine learning models to analyze LMS interactions logs and detect patterns associated with student engagement and academic performance.

The research design includes the following components:

1. **Data-Based Analysis**

Several Studies analyzed in this research rely on real- world datasets collected from e- learning platforms. For example:

- The **open University learning analytics datasets** includes **10 million student interaction events,** enabling researchers to model engagement patterns across multiple courses.
- The datasets used by Angeioplastis et al. [2] contains **nine semesters of academic performance records,** providing longitudinal data for predicative analysis

These datasets provide a reliable foundation for evaluating predictive models in educational settings.

2. **Feature Engineering Frameworks**

The research design emphasizes the importance of transforming raw LMS logs into meaningful behavioral indicators. Examples of engineered features include:

- Weekly login frequency
- Average time spent on learning resources
- Assignments Submission delays
- Participation in online discussions
- Learning activity consistency

These Features allow predictive models to identify associated with academic success or academic risk.

3. Predictive Modeling Techniques

The Research design examine various machine learning Algorithms commonly used to academic prediction systems, including:

- Random Forest
- Support vector machines
- Logistic Regression
- Neural Networks
- Deep Learning Architectures

These algorithms are evaluated based on their ability to analyze engineered features and predict student academic outcomes.

4. Explainability and Transparency

A critical aspect of the research design involves examining explainable AI techniques that help educators interpret predictive model Results feature importance ranking and scrap value analysis allow instructions to understand which behavioral indicators most strongly influence predictions. By combining these components, the research design provides a structured framework for analyzing predictive analytics systems and educational environments.

Data Collection

the data used in this research consists of secondary data collected from publicly available educational datasets and published academic research these lesser sets Originate from real e-learning platforms and educational institutes that record detailed information about student interactions with digital learning environments.

One of the most significant database used in the educational data mining research is the Open University Learning Analytics data set OULAD this This data sets contain information. About 32,593 students enrolled in multiple courses and more than 10 million elements interaction records. The datasets indicate variables such as:

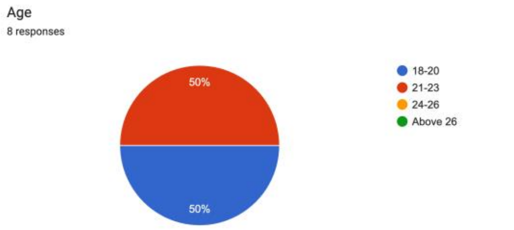
- Student demographic information
- Assessment scores
- Learning resources, access logs
- Forum discussion participation
- Assignment submission records

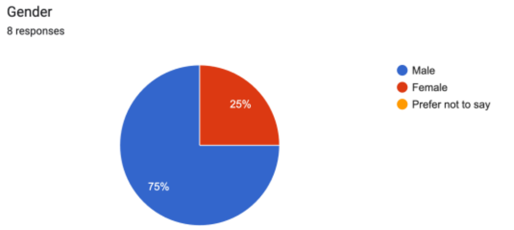
These data points allow researchers to analyze learning behavior patterns and identify indicators that influence academic success.

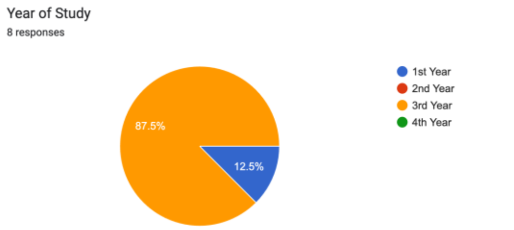
Another data set analyzed the previous studies involved moodle LMS interaction logs. Where researchers examined thousands of students activities recorded over the duration of online courses. These logs include timestamps for resource review, quiz attempts, and forum posts.

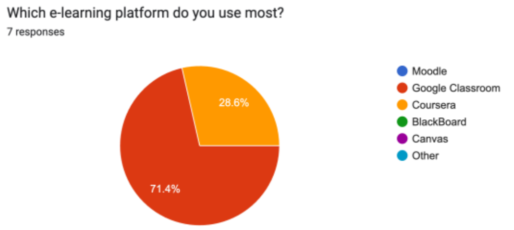
The study by Angioplasties et al. [2] You analyzed. Academic records from 450 students across 9 academic semesters, providing a longitudinal perspective on students learning behaviour pattern trends.

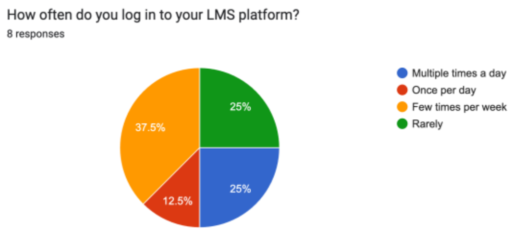
By reviewing the research studies that utilize these real world datasets, this study gains insights into how predictive analytics models operate in actual educational environments.

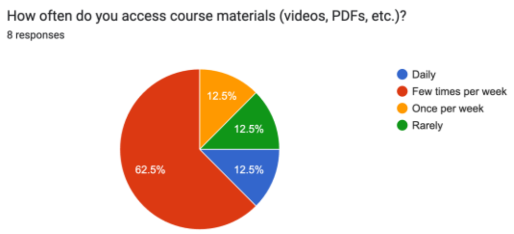


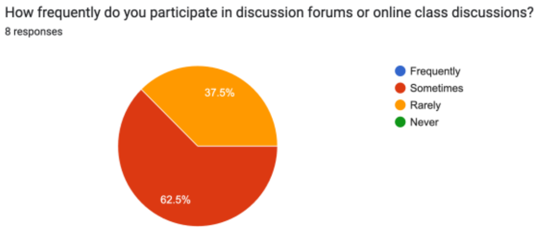


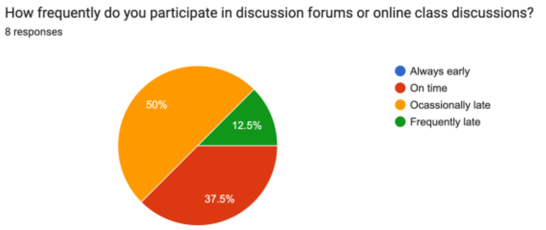


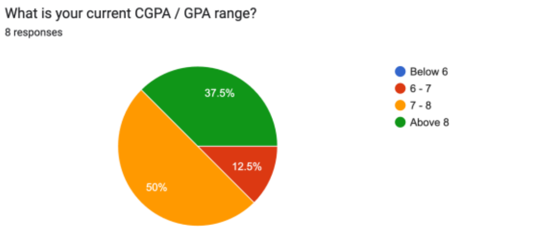


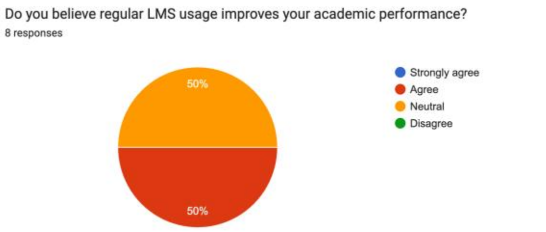












Data Analysis

The data analysis process used in this research follows A quantitative thematic analysis approach instead of performing new static modeling, the study analyzed findings from existing research papers to identify common pattern and. Related to feature engineering and Academic prediction models.

The analysis focus on several key aspects:

Behavioral feature identification.

His student consistently identifies several behavioral indicators that strongly correlate with academic performance. These include slogan frequency, time spent on learning materials, and assignment submission timing.

Research has shown that students who interact with LMS platforms more frequently tend to achieve higher academic performance compared to students who exhibit irregular learning behaviour.

Machine learning model performance.

Multiple The studies demonstrate that machine learning algorithm trained on engineered behavioral feature achieved significant higher prediction. Accuracy compared to model trained on raw LMS interaction log. I'm gonna search. Algorithms such as Random Forest and Support Vector Machine have shown strong performance in classifying student into performance categories.

Early warning systems.

Predictive models can be also be used to develop early warning systems that detect at risk student badgering the early week. Of course, by analyzing engagement pattern institution carry identify student who may Require additional academic support.

Explainable AI techniques.

Explainable AI methods such as SCHAP values help identify which feature contribute most strongly to model predictions. For example, feature importance in that analysis often reveals that Assignment Submission, punctuality and consistency. Engagement are among the strongest predictors of academic success.

Through this thematic analysis, the study highlights the critical role of feature engineering in enabling reliable, interpretable predictive models.

Scope of the Study

The Scope of this research focus on Predictive analytics systems used in digital learning environments, particularly those implemented through learning management systems. The study examined feature engineering techniques applied in higher education institutions, online learning platforms, and distance education systems.

the research covers predictive model java using LMS data sets From platform such as. Moodle, blackboard and Coursera. these systems generate behavioral interaction logs that provide valuable insights into student engagement and learning patterns.

The study also examined machine learning techniques commonly used in educational data mining research including random forests support vector machines and deep learning models.

however, this study primary focus on the role of feature engineering and improving prediction accuracy and interpretability rather than comparing specific algorithm in detail.

Limitation

despite providing valuable insights into academic prediction model this study has several limitations.

First, the research relies primarily on secondary data sources, including previous published research paper and publicly available datasets. As a result, I study does not include new experimental modeling or direct evaluation of predicting. Algorithms.

Second, the effectiveness of feature engineering techniques May vary depending on the specific educational context, including course structure. Subject area and student demographics.

Third, elements datasets may not capture all aspects of student learning behaviour. For example, offline study activities and informal learning interaction are not recorded in digital platform, which may limit the completeness of predictive models.

Finally, the rapid evolving nature of educational technique means that new data source and learning analytics tools continue to emerge, potentially influencing future research in this field.

Ethical Considerations

The use of predictive analytics in educational environments raises several ethical Everyone else's concerns related to fairness, transparency, and data privacy.

Student data collection from LMS platform often includes sensitive information such as academic performance record and behavioral interaction logs. Institution must ensure That such data is handled responsibly and in compliance with data protection regulations.
Feature engineering must also be designed carefully to avoid introducing biases that may disadvantage certain group of students. For example, model that rely heavily on logging. Frequency may in unintentionally penalize students who have limited access to Internet resources.

Explainable AI techniques play an important role in addressing these concerns by ensuring that the predictive models remain transparent and interpretable. Educators must be able to understand how predictions are generated to make. Information. Academic decisions.

By prioritizing fairness, accountability, and transparency, predictive analytics systems can understand positive educational outcomes while maintaining ethical standards.

Conclusion

The increasing adoption of e-learning platforms has created new opportunity for educational institutes to analyze student learning behavioral through digital interaction data. Machine learning techniques applied to LMS datasets enabled the development of predictive analytics systems that identify patterns associated with academics success and academic risk.

this research demonstrates that feature engineering is critical component in Transforming raw elements interaction logs into meaningful behavioral indicators that improve predictive modeling performance. Engineered features such as learning resistance management, engagement frequency, and assignment submission patterns allow machine learning models. To achieve higher accuracy while remaining interpretable for educators.

Studies using real world datasets including the Open University Learning Analytics datasets and. Moodle interaction logs shows that predictive models based on engineered featuring can effectively identify students who may require academic invention.

By integrating feature engineering with explainable AI techniques, educational institutes can develop predictive analytics systems that support data-driven decision making while maintaining transparency. Fairness and ethical responsibilities.

References

[1] S. boujmiraz H. Darhmaoui, and A.Drissi, “predicting student performance : A comprehensive review of machine learning and explainable AI approches “ computers and education : artificial intelligence, 2026

https://www.sciencedirect.com/science/article/pii/S2666920X26000093

[2] A. angeioplastis et al., “A data – driven approach using educational data mining to predict student performance,”

Computers, vol. 14, no. 3, 2025.

https://www.mdpi.com/2073-431X/14/3/83

[3] Yuan et al.,

“ integrating behavior analysis with machine learning to predict online learning performance,”

https://arxiv.org/abs/2406.11847

[4] G. Airlangga,

“predicting student performance using deep learning models,” 2024

https://journal.irpi.or.id/index.php/malcom/article/view/1668

[5] A. G. R. Sandeepa and S. Mohottala

“evaluation of machine learning models in student academic performance prediction,” 2025.

https://arxiv.org/abs/2506.08047

[6] “Predicting students’ performance using machine learning learning algorithms,” international journal of emerging technologies in learinternational journal of emerging technologies in learning, 2023

https://www.researchgate.net/publication/372491497_Predicting_Students%27_Performance_Using_Machine_Learning_Algorithms_A_Review