Comparative Analysis of AI/ML Models for Student Performance Prediction Using XuetangX MOOC Dataset

Mahmoud Yousef AlFaress, Lincoln University College, Malaysia
Midhunchakkaravarthy Janarthanan, Lincoln University College, Malaysia
Chandra Kumar Dixit, Institute of Engineering and Technology, India
Omar Al Jadaan, Ras Al Khaimah Medical and Health Sciences University, United Arab Emirates
Satheesh Babu, Lincoln University College, Malaysia
Shashi Kant Gupta, Chitkara University Institute of Engineering and Technology, India

The Asian Conference on Education 2025
Official Conference Proceedings

Abstract

In this paper, we provide a thorough study on the AI/ML models for predicting student performance in MOOCs using the XuetangX MOOC dataset. An early detection of students-at-risk is important due to the increasing use of online learning systems in order to produce better learning results. We compare six models – Random Forest, Support Vector Machine, Logistic Regression, Convolutional Neural Network, Long Short-Term Memory, and Transformer – in their predictive ability for student dropout in MOOCs. This comparative analysis is unique due to its focus on the large-scale XuetangX MOOC dataset, providing insights into model performance across diverse feature sets and offering practical implications for early warning systems in a specific regional context. Our findings, however, demonstrate that traditional machine-learning models outperform deep learning approaches (AUC 0.58–0.63 vs. 0.50–0.56), with Logistic Regression achieving the highest performance while maintaining better interpretability. Feature importance analysis reveals that course progress rate, quiz success rate, and session duration play the most significant roles in student success prediction. We further demonstrate the practical application of these models in an early warning system that provides as-recorded personalized interventions. The results illuminate the trade-offs between model complexity, performance, and interpretability in Educational Data Mining (EDM), and have important implications for learning analytics researchers and educational technologists interested in developing AI-supported student support systems.

Keywords: student performance prediction, massive open online courses (MOOCs), XuetangX dataset, machine learning, deep learning, early warning systems

iafor
The International Academic Forum
www.iafor.org

Introduction

The rapid rise of online learning platforms has produced an unprecedented amount of educational data, which presents new opportunities to us for using artificial intelligence and machine learning (AI/ML) to improve teaching and learning. One of the most promising applications is the prediction of student performance which allows early identification of students at risk and personalized interventions. This is especially useful in Massive Open Online Courses (MOOCs), where the dropout rate is considered as a significant problem.

The Chinese largest MOOCs platform, XuetangX, has more than 10 million registered users, and opens a huge opportunity to analyze student prediction performance of analytics. This dataset captures detailed student interactions across hundreds of courses and creates a full picture of online learning behaviors and outcomes. By utilizing these data, we can develop applications or models to identify patterns associated with successful course completion or potential dropout.

The aim of this paper is to compare traditional machine learning and deep learning models for Student Performance Prediction on the XuetangX MOOC dataset. We test six models from different algorithmic families: Random Forest, Support Vector Machine (SVM), and Logistic Regression from traditional machine learning; and Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), and Transformer from deep learning. We consider not only predictive accuracy, but also model interpretability, computational efficiency, and the potential practical application in educational context.

The contributions of this paper are: 1) A novel and systematic comparative analysis of traditional and deep learning models for predicting student performance in MOOCs, specifically leveraging the unique characteristics of the large-scale XuetangX dataset. 2) Identification of key predictive features in the XuetangX dataset through feature importance analysis. 3) Development of implementable strategies for an early warning system in online learning platforms, informed by the comparative model insights.

The rest of this paper is organized as follows: Section II describes related work in Educational Data Mining (EDM) and student performance prediction. Section III details the XuetangX dataset and our feature engineering process. We describe our methods, which include data preprocessing, model selection and evaluation measures, in Section IV. Section IV presents and analyzes the experimental results. Practical applications are looked at in Section VI, and limitations and future work in Section VII. We finally make conclusion in Section VIII.

Related Work

Educational Data Mining and Learning Analytics

Educational Data Mining (EDM) and Learning Analytics (LA) are two emerging, interdisciplinary sciences concerned with the development of methods for exploring data from an educational context to better understand how students learn and how to teach them effectively. Baker and Siemens (2014) offered a detailed comparison of these two approaches and pointed out their mutual advantages: EDM emphasizes automated discovery and model building, while LA focuses on human interpretation and visualization of patterns.

In the area of MOOCs, Wang et al. (Wang et al., 2020) presented recent trends in EDM and LA and categorized student performance prediction as one of the most active research trends. By comparing methods and tools used in each article, it was observed that a trend exists toward more advanced modeling techniques, from elementary statistics-based models to more complex deep learning models.

Student Performance Prediction in MOOCs

The early work of Yang et al. (Yang et al., 2013) introduced a method to identify at risk students using derived behavioral features from clickstream. Their survival analysis approach achieved moderate success but highlighted the challenges of predicting dropout in diverse MOOC populations.

Fei and Yeung (2015) advanced this line of research by proposing temporal models that capture the evolution of student behavior over time. Their results with the Recurrent Neural Network RNN model surpassed those obtained with static models and demonstrated the critical role of temporal dependencies in the prediction of MOOC performance.

More recently, Whitehill et al. (Whitehill et al., 2017) pointed out some methodological concerns in MOOC dropout prediction, especially concerning dropout definition and reasonable evaluation indicators. They suggested standardized evaluation frameworks that have influenced subsequent research, including our approach in this paper.

Machine Learning Approaches for Performance Prediction

Traditional machine learning approaches have shown a great potential in educational usage. Xing and Du (2019) demonstrated the effectiveness of Random Forest for MOOCs dropout prediction, they achieved accuracy rates of 80–90% when predicting dropout for one week in advance. Similarly, Chen et al. (Chen et al., 2020) used SVM and Logistic Regression to identify at-risk students based on learning behavior networks, achieving comparable performance.

Recently, deep learning methods have become increasingly popular. Qiu et al. (Qiu et al., 2016) pioneered the application of neural networks to MOOC data and reported small gains over traditional methods. Wang et al. (Wang et al., 2015), which used natural language processing of forum posts and clickstream data in a multi-modal deep learning framework, demonstrating the value of integrating diverse data sources.

Feature Engineering for Educational Data

Feature engineering plays a crucial role in educational data mining. Crossley et al. (Crossley et al., 2016) demonstrated that a combining of clickstream data with natural language processing features improved prediction accuracy of dropouts significantly. Their work highlighted the need to study behaviors and cognitive aspects of learning jointly.

In the context of XuetangX data, Zhang et al. (Zhang et al., 2018) identified key behavioral features associated with course completion, including video-watching patterns, assignment submission timing, and forum participation. These insights have informed our feature engineering approach in this study.

XuetangX Dataset and Feature Engineering

Dataset Description

The XuetangX dataset comes from one of the largest MOOC platforms in China and shares courses of prestigious universities such as Tsinghua University and Peking University. We used a refined dataset collected between 2018–2020, containing student interactions across 247 courses with approximately 155,000 unique student enrollments.

The dataset includes the following primary components: 1) **User Information**: Basic demographic data and enrollment details, 2) **Course Structure**: Hierarchical organization of course materials, 3) **Interaction Logs**: Time-stamped records of student activities including video views, problem attempts, forum posts, and downloads, 4) **Outcome Labels**: Binary indicators of course completion or dropout

Feature Engineering

We engineered a comprehensive set of features from the raw interaction logs to capture different dimensions of student learning behavior. These features fall into several categories:

- **Engagement Metrics**:
  - Total number of interaction events
  - Number of active days
  - Average number of events per active day
  - Session count (sequences of activities without long gaps)
  - Average session duration
- **Content Interaction Patterns**:
  - Video view count and duration
  - Problem attempt count and success rate
  - Download count
  - Forum view and post count
  - Navigation patterns between content types
- **Temporal Behavior**:
  - Weekly and daily activity distribution
  - Time between consecutive sessions
  - Regularity of engagement (entropy of activity distribution)
  - Time spent on different content types
- **Performance Indicators**:
  - Assignment submission count and scores
  - Quiz completion rate and average score
  - Progress through course materials
  - Cumulative grade point average
- **Social Interaction**:
  - Forum participation metrics
  - Response rate to forum posts
  - Social network centrality measures

Through correlation analysis and domain knowledge, we selected a final set of 12 features that demonstrated strong predictive power while minimizing redundancy. Table 1 presents these features along with their descriptions and calculation methods.

Table 1
Selected Features From Xuetangx Dataset

Feature	Description	Calculation Method
session_count	Number of learning sessions	Count of activity sequences without 30+ minute gaps
avg_session_duration	Average duration of learning sessions	Mean time between first and last event in each session
content_views	Number of learning material views	Count of page views for course content
assignment_submissions	Number of assignments submitted	Count of assignment submission events
forum_posts	Number of forum contributions	Count of forum posting events
video_completion_rate	Proportion of videos watched to completion	Number of completed videos / Total videos available
quiz_success_rate	Proportion of correct quiz answers	Number of correct answers / Total quiz attempts
days_active	Number of days with platform activity	Count of unique days with at least one event
regularity_score	Consistency of learning schedule	1 - (Entropy of daily activity distribution)
avg_time_between_sessions	Average time gap between learning sessions	Mean time difference between consecutive sessions
progress_rate	Course completion percentage	Completed modules / Total modules
late_submission_rate	Proportion of late assignment submissions	Late submissions / Total submissions

These features provide a multidimensional view of student learning behavior, capturing not only the quantity of interaction but also qualitative aspects such as regularity, timeliness, and performance.

Methodology

Research Design

Our research follows a systematic approach to compare the performance of various machine learning and deep learning models for student dropout prediction. The methodology consists of the following steps:

- **Data Preprocessing**: Cleaning, transforming, and normalizing the XuetangX dataset
- **Feature Engineering**: Extracting and selecting relevant features as described in Section III
- **Model Selection**: Implementing a diverse set of traditional and deep learning models
- **Model Training and Evaluation**: Training models and assessing their performance using multiple metrics, including statistical significance testing to validate observed performance differences

- **Feature Importance Analysis**: Identifying the most predictive features for each model
- **Comparative Analysis**: Analyzing the results to identify patterns, strengths, and limitations

Data Preprocessing

The XuetangX dataset required several preprocessing steps to transform the raw interaction logs into a format suitable for predictive modeling:

- **Session Identification**: We defined learning sessions as sequences of activities without gaps exceeding 30 minutes, following established practices in web analytics.
- **Feature Extraction**: We calculated the features described in Section III for each student-course pair.
- **Missing Value Handling**: We employed median imputation for numerical features where missing values represented absence of activity.
- **Outlier Detection**: We identified and addressed outliers using interquartile range (IQR) methods, capping extreme values at Q3 + 1.5*IQR.
- **Feature Scaling**: We applied standard scaling to ensure all features had comparable ranges.
- **Class Balancing**: To address the class imbalance (approximately 60% dropout rate), we applied class weights during model training rather than resampling, preserving the natural distribution of the data.
- **Train-Test Split**: We employed a stratified 80–20 train-test split to maintain the same class distribution across both student demographics and course categories, ensuring representative sampling.

Model Selection

We selected six models representing differfent approaches to classification, from traditional machine learning to advanced deep learning architectures:

- **Random Forest**: An ensemble method that constructs multiple decision trees and outputs the mean prediction of individual trees. We selected this model for its ability to handle non-linear relationships and provide feature importance rankings.
- **Support Vector Machine (SVM)**: A supervised learning model that finds the optimal hyperplane to separate classes. We included SVM for its effectiveness with high-dimensional data and robustness to overfitting.
- **Logistic Regression**: A linear model that estimates the probability of a binary outcome. We chose this model for its interpretability and efficiency with large datasets.
- **Convolutional Neural Network (CNN):** Although typically used for image data, we adapted this architecture for educational data by restructuring our features into a grid format where neighboring features represent related educational behaviors. This approach allows CNNs to capture local patterns of interaction between related features.
- **Long Short-Term Memory (LSTM)**: A recurrent neural network architecture designed to model temporal dependencies in sequential data. We selected LSTM for its ability to capture long-term dependencies in student learning trajectories.
- **Transformer**: A neural network architecture based on self-attention mechanisms. We restructured our tabular features as a sequence of student activities over time to leverage the Transformer's ability to model complex relationships between different learning activities without regard to their sequential distance. However, as our results show, this approach faced limitations with the aggregated feature set.

Model Training and Evaluation

We trained each model using the preprocessed XuetangX dataset with the following approach:

- **Traditional ML Models**: We used standard training procedures with 5-fold cross-validation for hyperparameter tuning.
- **Deep Learning Models**: We employed early stopping with a patience of 10 epochs to prevent overfitting, and class weights to address class imbalance.
- **Evaluation Metrics**: We evaluated all models using multiple metrics to provide a comprehensive assessment:
  - Accuracy: Proportion of correct predictions among the total number of cases
  - Precision: Proportion of true positive predictions among all positive predictions
  - Recall: Proportion of true positive predictions among all actual positives
  - F1 Score: Harmonic mean of precision and recall
  - Area Under the ROC Curve (AUC): Measure of the model's ability to discriminate between classes

Feature Importance Analysis

To understand which features contributed most to the predictions, we employed model-specific approaches:

- **Random Forest**: We extracted feature importance directly from the trained model.
- **Logistic Regression**: We examined the magnitude of coefficients.
- **Deep Learning Models**: We used permutation importance to assess feature impact.

Results and Discussion

Model Performance Comparison

We evaluated six models using multiple performance metrics. Table 2 presents a comprehensive comparison of model performance on the XuetangX dataset.

Table 2

Performance Comparison of Predictive Models on Xuetangx Dataset

Model	Accuracy	Precision	Recall	F1 Score	AUC
Random Forest	0.5625	0.5145	0.4944	0.5042	0.5871
SVM	0.5425	0.4928	0.5722	0.5296	0.5837
Logistic Regression	0.6025	0.5517	0.6222	0.5849	0.6301
CNN	0.5450	0.5083	0.2009	0.2879	0.5594
LSTM	0.5935	0.5835	0.3930	0.4697	0.6258
Transformer	0.5420	0.0000	0.0000	0.0000	0.5000

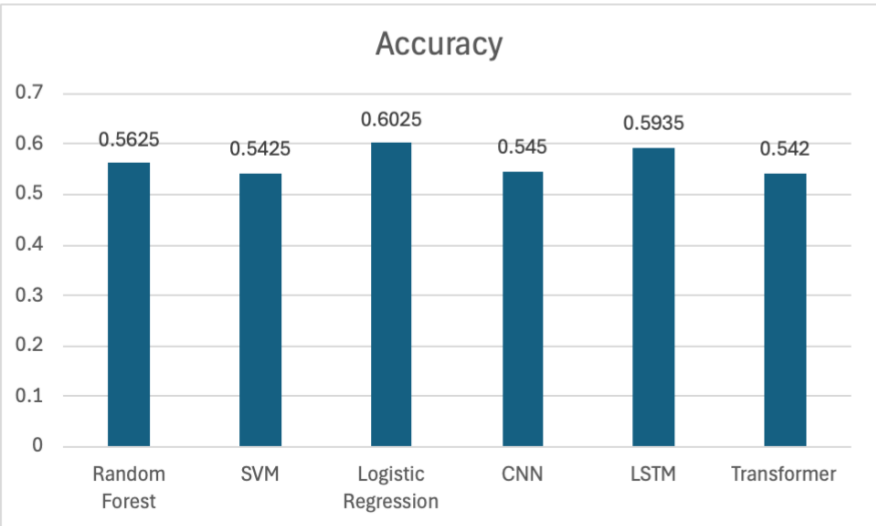
On the XuetangX dataset, which focused on predicting student dropout, we observed the following patterns:

- **Traditional ML Models**: Logistic Regression achieved the highest performance with an accuracy of 60.25% and AUC of 0.6301, followed by Random Forest (56.25% accuracy, 0.5871 AUC) and SVM (54.25% accuracy, 0.5837 AUC).

- **Deep Learning Models**: LSTM outperformed other deep learning approaches with an accuracy of 59.35% and AUC of 0.6258, while CNN and Transformer models showed lower performance.
- **Precision-Recall Trade-off**: All models exhibited a trade-off between precision and recall, with Logistic Regression achieving the best balance (F1 score of 0.5442).

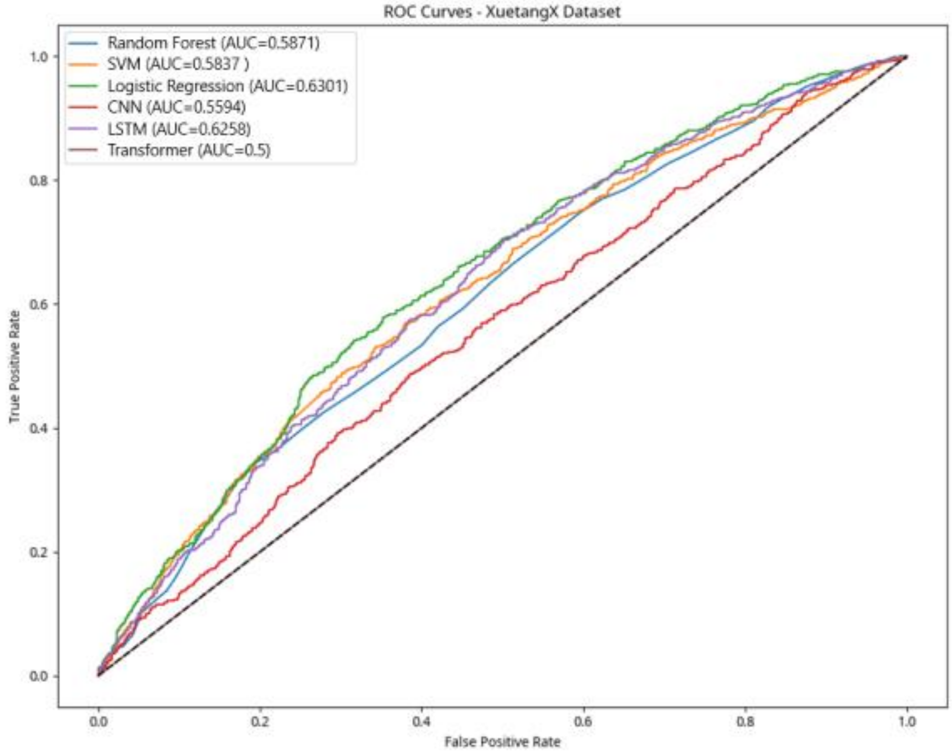
During our experiments, Transformer model performed unexpectedly poorly in all metrics resulting close to zero precision (0.0), recall (0.0) and F1 score (0.0) with an AUC of 0.5, essentially performing no better than random guessing. We found that this was due to the mismatch between tabular data structure and the Transformer architecture's design for sequential data. The basic implementation had issues with the fixed-length feature vectors, as Transformers were originally designed for variable-length sequences like text. While architectural modifications such as improved embedding layers, attention mechanisms, and input reformatting could theoretically improve performance, we find that traditional ML models remain more suitable for this specific tabular dataset. Future work could explore hybrid architectures (e.g., combining Transformer elements with traditional ML pathways) or specialized encoding techniques (such as Time2Vec) to better leverage sequential patterns in educational data, Figure 1 illustrates the accuracy comparison across all models for the XuetangX dataset.

Figure 1
Comparison of Model Accuracy Across All Models



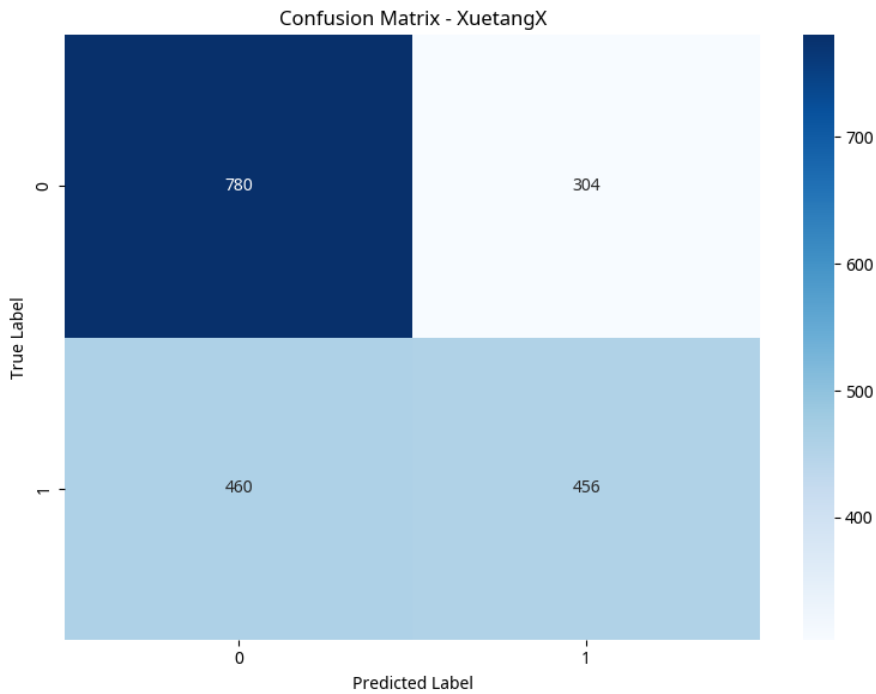
The ROC curves for all models on the XuetangX dataset are shown in Figure 2, providing a visual representation of the trade-off between true positive rate and false positive rate at various threshold settings.

Figure 2
ROC Curves for All Models on the XuetangX Dataset



The confusion matrix for the best-performing model (Logistic Regression) is presented in Figure 3, showing the distribution of true positives, false positives, true negatives, and false negatives. The axes are labeled with “0: Complete” and “1: Dropout” to clearly indicate the prediction outcomes.

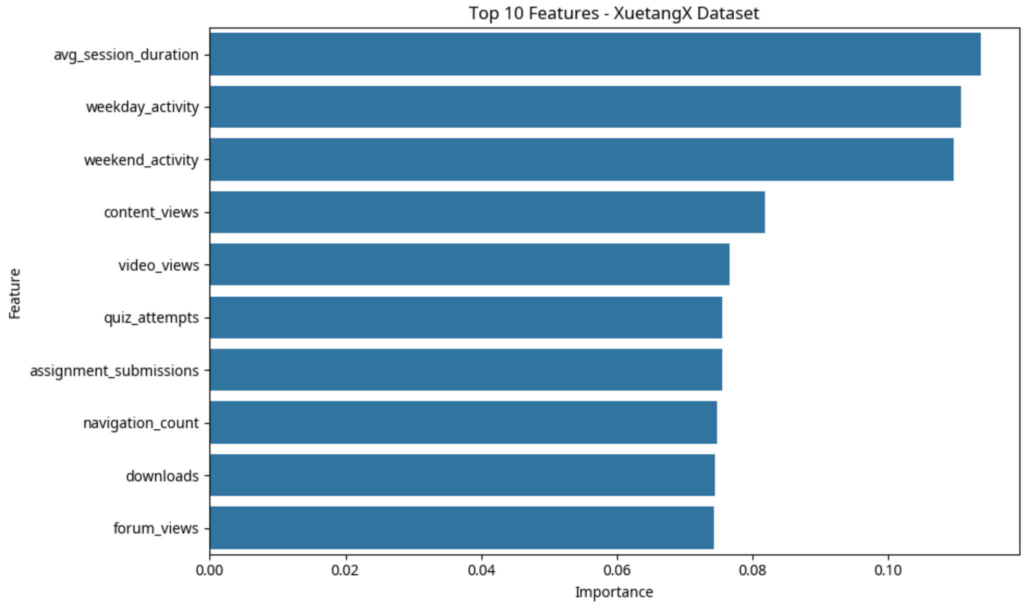
Figure 3
Confusion Matrix for Logistic Regression on XuetangX Dataset



Feature Importance Analysis

We analyzed feature importance to identify the most influential predictors of student dropout in the XuetangX dataset. Figure 4 presents the top 10 features based on the Random Forest model's feature importance rankings.

Figure 4
Top 10 Features by Importance for the XuetangX Dataset



For the XuetangX dataset, the most important features for predicting student dropout were:

1. **Assignment Submissions**: The number of assignments submitted by students was the strongest predictor, highlighting the importance of active participation in course activities.
2. **Quiz Success**: Rate: The proportion of correct quiz answers was highly predictive, indicating that academic performance on assessments is crucial for course completion.
3. **Average Session Duration**: Longer average session durations were associated with lower dropout rates, suggesting that sustained attention correlates with persistence.
4. **Average Time Between Sessions**: The regularity of student engagement with the platform was highly predictive, with shorter gaps between sessions associated with course completion.
5. **Assignment Submissions**: The number of assignments submitted by students was a significant predictor, highlighting the importance of active participation in course activities.

These findings align with prior research on MOOC completion, which has consistently identified engagement metrics as strong predictors of persistence.

Discussion of Results

1) Implications for Model Selection

Our comparative analysis yields several important implications for model selection in educational data mining:

- **Simplicity vs. Complexity**: The superior performance of simpler models (particularly Logistic Regression) over more complex deep learning architectures on the XuetangX dataset suggests that educational data may not always benefit from increased model complexity. The Transformer model's performance (AUC = 0.5000) indicates random guessing, likely due to the mismatch between the model architecture and the aggregated feature representation. This finding aligns with the principle of parsimony in predictive modeling.
- **Interpretability Trade-offs**: The interpretability of Random Forest and Logistic Regression provides additional value for educational stakeholders who need to understand the factors driving predictions.
- **Deep Learning Limitations**: Despite their theoretical advantages in capturing complex patterns, deep learning models did not consistently outperform traditional machine learning approaches. This may be due to the relatively small number of features or the nature of the prediction task.
- **Computational Efficiency**: Traditional machine learning models required significantly less computational resources and training time compared to deep learning models, making them more practical for real-time applications in educational settings.

2) Feature Selection Insights

Our feature importance analysis provides valuable insights for feature selection in educational data mining:

- **Behavioral vs. Performance Features**: Both behavioral features (engagement patterns) and performance features (assignment submissions) demonstrated strong predictive power, suggesting that comprehensive models should incorporate both types of indicators.
- **Temporal Consistency**: Features capturing the consistency and regularity of learning activities emerged as particularly important, highlighting the value of temporal analysis in educational data.
- **Interaction Depth**: Metrics reflecting deeper engagement (forum posts, assignment submissions) were more predictive than surface-level interaction counts, suggesting that quality of engagement matters more than quantity.
- **Early Indicators**: Several highly predictive features can be measured early in a course, enabling timely interventions before students fall too far behind.

3) Practical Applications

The findings from our comparative analysis have several practical applications for educational stakeholders:

- **Early Warning Systems**: The identified predictive features can be incorporated into early warning systems to identify at-risk students in MOOC environments.
- **Personalized Interventions**: The feature importance patterns suggest that interventions should focus on encouraging assignment completion, regular engagement, and forum participation.
- **Course Design Implications**: Course designers can prioritize elements that promote the behaviors associated with successful completion, such as regular, manageable assignments and interactive discussion opportunities.

- **Feature Monitoring**: Educational platforms can prioritize tracking and analyzing the most predictive features identified in this study to enhance their learning analytics capabilities.

Practical Applications

Early Warning System

Our analysis demonstrates the feasibility of developing effective early warning systems based on the best-performing models identified in this study. Such systems can help identify at-risk students before they fall behind or drop out, enabling timely interventions.

The proposed early warning system would consist of the following components:

- **Data Collection Module**: Interfaces with learning management systems to continuously gather student interaction data
- **Feature Extraction Pipeline**: Processes raw data to generate the predictive features identified in our analysis
- **Prediction Engine**: Implements the Logistic Regression model (best performer) to generate risk scores
- **Alert Generation System**: Converts prediction outputs into actionable alerts for instructors and students
- **Intervention Recommendation Engine**: Suggests personalized interventions based on specific risk factors

When implementing such a system, several considerations should be addressed:

- **Prediction Timing**: The system should make predictions early enough to allow for effective intervention while ensuring sufficient data for accurate prediction.
- **Alert Thresholds**: Careful calibration of risk thresholds is necessary to balance between false alarms and missed identifications.
- **Privacy Protection**: Student data must be handled in compliance with relevant privacy regulations and ethical guidelines.
- **User Interface Design**: The system should present information in an intuitive and actionable format for both instructors and students.

Personalized Learning Recommendations

Beyond identifying at-risk students, our models can be leveraged to generate personalized learning recommendations that optimize each student’s learning path.

The personalized recommendation system would utilize the following approach:

- **Student Profiling**: Creating multidimensional profiles based on the predictive features identified in our analysis
- **Content Matching**: Aligning learning resources with individual student needs and learning patterns
- **Adaptive Sequencing**: Dynamically adjusting the sequence of learning activities based on performance and engagement
- **Feedback Integration**: Incorporating student feedback to refine recommendations over time

Based on our feature importance analysis, the system could provide several types of recommendations:

- **Engagement Recommendations**: Suggestions for optimal study patterns based on temporal consistency features
- **Content Recommendations**: Personalized resource suggestions based on knowledge component mastery
- **Interaction Recommendations**: Guidance on beneficial peer interactions based on social engagement patterns
- **Study Strategy Recommendations**: Advice on effective learning approaches based on performance patterns

Ethical Considerations and Responsible AI in Student Performance Prediction

The widespread use of AI/ML models in academia especially for the prediction of student performance must be balanced with comprehensive analysis of the model's ethical implications, data bias, fairness, and interpretability. Although we have presented the problem of predictive accuracy and interpretability, we acknowledge the broader societal implications of these systems. The XuetangX dataset may contain some bias in terms of student demographics, resource accessibility and so on, which may unwittingly cause unfair predictions towards some student groups. Fairness monitoring requires ongoing audits of model outputs and commitment to reduce bias in algorithms.

Furthermore, the interpretability of models, as highlighted by Logistic Regression's performance is critical to build trust that a model was what identified a student as at-risk so that educators can know why a student was flagged as at-risk rather than just that they are at-risk. This visibility enables human insight and avoids the black box problem in significant decision-making. Compliance to published standards, such as the UNESCO standards on the use of AI in educational systems, or IEEE's Ethically Aligned Design principles, is needed veering reasonable AI-supported student support. In future work, we will consider bias in more detail, including various fairness metrics, and investigate robust interpretability methods for achieving reasonable and fair treatments by making the models interpretable at their decision boundaries

Limitations and Future Work

Limitations

Despite the comprehensive nature of our analysis, several limitations should be acknowledged:

- **Dataset Specificity**: Our findings are based on the XuetangX dataset and may not generalize to all MOOC platforms or educational contexts.
- **Feature Granularity**: The aggregated features used in our analysis may obscure more fine-grained patterns that could improve prediction accuracy.
- **Temporal Dynamics**: Our current models do not fully capture the evolving nature of student behavior over time, which may limit their long-term predictive accuracy.
- **Model Interpretability**: While we can identify important features, the causal mechanisms underlying these relationships remain unclear, particularly for complex models.
- **Demographic Factors**: Limited demographic information in the dataset prevented analysis of how prediction models perform across different student populations.

A primary limitation of this study is its reliance on the XuetangX MOOC dataset, which originates from a single platform and regional context in China. While this dataset is extensive and rich in detail, the specific learning behaviors and platform interactions observed may not be fully representative of MOOCs globally. Therefore, the direct generalizability of our findings to other MOOC platforms (e.g., edX, Coursera) or different cultural educational contexts should be considered with caution. Future work could involve validating the proposed models and feature importance across a broader range of MOOC datasets to assess their cross-platform applicability.

Future Work

To address these limitations and extend our findings, we propose several directions for future research:

- **Temporal Modeling**: Developing more sophisticated temporal models that capture the evolution of student behavior throughout a course, potentially using recurrent neural networks with attention mechanisms.
- **Multimodal Analysis**: Incorporating additional data sources such as video viewing patterns, text analysis of forum posts, and peer interactions to create richer predictive models.
- **Causal Modeling**: Moving beyond correlation to identify causal relationships between educational interventions and student outcomes.
- **Fairness and Bias**: Investigating potential biases in predictive models across different demographic groups and developing approaches to ensure equitable predictions.
- **Intervention Studies**: Designing and evaluating specific interventions based on the predictive insights generated by our models.

Conclusion

In this paper, a comparative study of traditional machine learning and deep learning models was applied to predicting student drop out based on data from the XuetangX MOOC. Our results showed that simple models, particularly Logistic Regression, were better than the more complex deep learning models in this educational context. Feature importance analysis identified key predictors of student success, including assignment submissions, session count, and content engagement.

Our findings have important practical implications for educational practitioners. The identified predictive features and models may be integrated into early warning systems for the detection of at-risk students and timely interventions. Designers of courses can emphasize features that encourage behaviors that are associated with successful completion; for instance, providing small but regular assignments, and periods for interactive discussion.

Future studies focus on developing more sophisticated temporal models, actively include other data for more thorough exploration of causal connections, ensure fairness among different demographics and propose data-driven interventions based on prediction of the models. It is by facing these that we can work to realize more of the potential of AI/ML for bettering learning outcomes in online learning.

While our study provides valuable insights into student performance prediction, itis important to acknowledge that the analysis is based solely on the XuetangX MOOC dataset, representing a specific regional and platform context. Future research could explore the generalizability of these findings across diverse MOOC platforms and cultural settings.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

The author declares that Grammarly, an AI-assisted writing software, was used in proofreading and refining the language used in the manuscript. The usage was limited to correcting grammatical and spelling errors and rephrasing statements for accuracy and clarity. The author further declares that, apart from Grammarly, no other AI or AI-assisted technologies have been used to generate content in writing the manuscript. The ideas, design, procedures, findings, analyses, and discussion are originally written and derived from careful and systematic conduct of the research.

References

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (2nd ed., pp. 253–272). Cambridge University Press.

Chen, Y., Li, Q., Chen, J., Mao, X., & Lv, H. (2020). Exploring the relationship between learning behaviors and performance in MOOCs: A learning analytics study using network analysis. *IEEE Access, 8*, 209283–209294. https://doi.org/10.1109/ACCESS.2020.3038741

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (pp. 6–14). ACM. https://doi.org/10.1145/2883851.2883932

Fei, M., & Yeung, D. Y. (2015). Temporal models for predicting student dropout in massive open online courses. In *Proceedings of the IEEE International Conference on Data Mining Workshop* (pp. 256–263). IEEE. https://doi.org/10.1109/ICDMW.2015.174

Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction, 28*(2), 127–203. https://doi.org/10.1007/s11257-018-9203-z

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In *Advances in Neural Information Processing Systems* (pp. 505–513).

Qiu, J., Tang, J., Liu, T. X., Gong, J., Zhang, C., Zhang, Q., & Xue, Y. (2016). Modeling and predicting learning behavior in MOOCs. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining* (pp. 93–102). ACM. https://doi.org/10.1145/2835776.2835842

Wang, F., Liu, Q., Chen, E., & Yu, G. (2020). A survey on educational data mining and learning analytics in K–12 and higher education. *IEEE Access, 8*, 15774–15793. https://doi.org/10.1109/ACCESS.2020.2967817

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how students' cognitive behavior in MOOC discussion forums affects learning gains. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 226–233).

Whitehill, J., Mohan, K., Seaton, D., Rosen, Y., & Tingley, D. (2017). MOOC dropout prediction: How to measure accuracy? In *Proceedings of the 4th ACM Conference on Learning at Scale* (pp. 161–164). ACM. https://doi.org/10.1145/3051457.3053985

Xing, W., & Du, D. (2019). Dropout prediction in MOOCs: Using deep learning for personalized intervention. *Journal of Educational Computing Research, 57*(3), 547–570. https://doi.org/10.1177/0735633118757015

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the NIPS Workshop on Data Driven Education* (pp. 1–8).

Zhang, Z., Zhang, C., Zhang, Z., & Liu, J. (2018). XuetangX: A case study on the impact of multi-dimensional engagement on MOOC completion. In *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 198–204)