Jiang, B. et al. (Eds.) (2025). Proceedings of the 33rd International Conference on Computers in Education. Asia-Pacific Society for Computers in Education

Detecting Cheating in Online Education: A Machine Learning Approach with Label Noise Analysis

Pham-Duc THO[a], Nguyen-Anh TU[b] & Bui-Thuy DUONG[c]
[a]*Faculty of Applied Sciences, Vietnam National University Hanoi, Vietnam*
[b]*Hanoi University of Industry, Vietnam*
[c]*Vietnam National University Hanoi, Vietnam*
*nguyenanhtu10203@gmail.com

Abstract: The growth of online education has created new challenges for maintaining academic integrity. This study perfoms a comparative of three machine learning models for detecting cheating behavior in online learning platforms. We used the Junyi Academy Online Learning Dataset containing 12,537 student interactions from May 2018 to June 2019. Three machine learning models were evaluated: XGBoost, LightGBM, Random Forest, and AdaBoost. We developed a systematic labeling methodology based on three established principles including fast completion times, performance improvements, and group collaboration patterns. To test model robustness, we performed random label noise at levels of 10%, 20%, 30%, 40%, and 50% to simulate real-world labeling errors. Results demonstrate that Random Forest achieves the highest cheating detection capability with 100% recall and 97% accuracy under noise conditions below 10%, maintaining 87% accuracy at 30% and 75% accuracy at 40% noise levels. These findings suggest that the bagging ensemble learning method, specifically Random Forest, is effective for automated cheating detection in education and shows superior noise resistance compared to gradient boosting approaches.

Keywords: Cheating detection, machine learning, online learning, academic integrity, label noise

1. Introduction

The rapid shift to online education has not only expanded access to learning but also made it more difficult to maintain academic integrity. without in-person supervision, online education creates new opportunities for cheating, while traditional proctoring methods often fail in large-scale online exams. Cheating is becoming more sophisticated, requiring better cheating detection systems. The use of advanced technology, peer collaboration and other diverse forms of cheating makes it harder to detect. As cited in in the study of McDonnell and Tantong (2023) about 60% of faculty believe that cheating is more prevalent online, which threatens the reputation of the institution and devalues the degree. Detecting cheating in online learning is complex for several reasons. Large-scale platforms make manual monitoring impractical, the variety and subtlety of digital cheating require advanced detection tools, and the risk of false positives can have serious consequences for students. In addition, deciding what counts as cheating often involves subjective judgment, which creates uncertainty in labeling data for model training. Machine learning offers promising solutions by automatically identifying patterns of behavior that may indicate cheating. However, its application faces specific challenges, such as dealing with noisy or incomplete data. In this research, we conducted a comparative study of three Machine Learning models robustness under label noise for cheating detection in online education. With our labeling rules, the results show that ensemble method performs better than gradient boosting models in this case. Random Forest is better model to detect cheating under noise conditions.

2. Literature Review

2.1 Academic Dishonesty in Online Education

Academic dishonesty is defined as dishonest behavior in doing an assignment, the result of which does not reflect the true level and knowledge of the person doing it. The foundational work of McCabe and Trevino (1993) established that academic integrity violations happens in every type of education, but online education presents unique challenges and opportunities for both students who cheat and for systems that try to catch them. The manifestations of online cheating are diverse and continuously evolving. We can easily see some basic forms such as copying answers or working together without permission, etc. For now, it has been enhanced by sophisticated digital strategies including using more than one device, getting live help from others, and taking advantage of weaknesses in software (Watson & Sottile, 2010).

Cheating in online learning happens for many reasons. Some come from outside pressure, such as parents expecting students to reach goals that are too high. Others come from the students themselves, especially in systems that focus on test scores. In these cases, students may cheat to get high marks or to avoid feeling less capable than classmates (Jalilzadeh et al., 2024). The reasons for these fraudulent thoughts come from reduced supervision, easier access to external resources, and perceived anonymity of digital platforms.

The consequences of online cheating extend beyond individual academic outcomes. For schools, it damages the reputation of programs and reduces the their degrees valuable in job applications. For society, it means some graduates may lack the skills they need for their jobs, which can harm workplace performance and weaken public trust in educational systems (Sozon et al., 2024).

2.2 Machine Learning in Cheating Detection Applications

The application of machine learning to educational data has grown significantly in recent years. This growth is due to the availability of large-scale educational datasets and advances in algorithmic sophistication. Educational data mining has addressed various challenges, such as predicting student performance, detecting students at risk of dropping out, and giving personalized learning advice. Several studies have applied machine learning techniques to cheating detection with varying degrees of success.

Garg and Goel (2025) introduced a machine learning method to detect “in-parallel collusion” in online exams. This type of cheating happens when students work together in real time. The researchers built a special quiz tool to record detailed click data. From this data, they created seven measures of behavioral similarity. These measures were used to train a Random Forest model, which achieved 98.8% accuracy in finding collusion. They suggest that adding different types of behavioral features to assessment systems can help make online tests fairer and more honest. Another research also introduced an intelligent assessment module for online lab exams that detects cheating by analyzing students’ mouse movements (Hassan Hosny et al., 2022). In this work, LightGBM was one of several machine learning algorithms evaluated. This model achieved the best results among all tested algorithms, with 90% accuracy, 88% precision, and the degree separation is 95%. The study highlights that LightGBM is particularly effective for this task due to its ability to handle complex feature interactions and large datasets efficiently.

Edrem and Karabatak (2025) proposed a method to classify and detect cheating patterns in online exams. Their study used data from 129 exams, with the target variable based on expert ratings of response time. After preprocessing, they tested three tasks: predicting the target value, binary classification (cheated or not), and three-class classification (cheated, uncertain, not cheated). The XGBoost model achieved the highest accuracy of 97.7% for students who cheated due to unethical behavior and also had the best performance on all other metrics.

3. Methodology

3.1 Dataset

In our research, we use only the Junyi Academy Online Learning Dataset from Kaggle. This dataset comes from a well-known Taiwanese online learning platform, which contains over 16 million problem attempts by 72,630 students in the 2018–2019 school year The data are stored in three tables: Info_UserData (student information), Info_Content (exercise information), and Log_Problem (detailed attempt logs). Junyi Academy is a valuable resource for detecting potential cheating because it provides interactive exercises, learning logs, and multiple attempts data. For the analysis, only attempts made between May 1 and June 11, 2019, were used. The chosen time window falls within the regular school term, when student activity is stable. This helps us avoid biases that may occur during long holidays, when learning behaviors are atypical. Besides, we selected only the fields related to features describing the problem and its difficulty, the learner's performance such as correctness, time spent, attempts and hints as well as the learning context like self-practice or classroom setting, while removing all personal identifiers to protect privacy. And then we combine them into one final table for cheating prediction.

3.2 Data Processing

We used the Junyi Academy Online Learning Activity Dataset from Kaggle, which contains user information, course content, and detailed test logs. After selecting the needed features, we cleaned the data by handling missing difficulty values and converting timestamps for labeling. Cheating labels were assigned using clear rules, with "0" for not cheating and "1" for cheating. The data was split 80:20 into training and test sets, and varying levels of label noise (10%, 20%, 30%, 40%, 50%) were added to the training set to test model robustness. We trained three models, including XGBoost, LightGBM, and Random Forest and evaluated them using confusion matrices and ROC-AUC. Finally, we compared results across scenarios to find the model with the best generalization and resistance to noise.

3.3 Labeling Rules

Our rules are established based on references to prior studies as well as our own subjective judgment. Related to some previous approaches (such as the "Score-Time-Ratio" (Xiao et al., 2022), timeline analysis (Du et al., 2022), and checks for abnormal score jumps or outlier performance (Kamalov et al., 2021)) we decided to build a set of 4 rules as follows:

Rule 1. Abnormal completion times: Students who answer significantly faster than their typical response time, get correct answers on first try, and use no hints may suggest prior access to questions or external assistance.

Rule 2. Sudden significant performance improvement: A sudden increase in performance, especially after previous low results, indicates unusual behavior possibly due to external access to answers.

Rule 3. Duplicate correct answers at identical time: When more than three different users answer the same non-easy question correctly at the same time, it suggests answer sharing or collaboration.

Rule 4. Unusual user behavior with high correct rate: Users with high rates of fast and correct answers across multiple consecutive questions without using hints indicate abnormal behavior.

4. Results

4.1 Performance Under Increasing Noise Levels

All models in the research achieved high accuracy under low-noise conditions (10% - 20%), supporting the feasibility of automated cheating detection using Machine Learning. As shown in Table 1, noise levels under 10% resulted in 99% accuracy for Random Forest and, while XGBoost achieved 97% and LightGBM achieved 96%. When noise increased, Random Forest demonstrated the strongest robustness by keeping an accuracy of over 75% even at 40% noise. XGBoost and LightGMB just achive 64% and 63% respectively.

Table 1. *Model Accuracy under Different Noise Levels*

Noise Level	XGBoost	LightGBM	Random Forest
10%	0.97	0.96	0.99
20%	0.87	0.87	0.97
30%	0.77	0.75	0.87
40%	0.64	0.63	0.75
50%	0.49	0.49	0.47

4.2 Cheating Detection Capability

Table 2 presents recall scores for detecting cheating cases (positive class). Random Forest maintained consistent recall performance (98-100%) across 10-40% noise levels, demonstrating exceptional capability for detecting actual cheating instances. Although the accuracy is equal at 10% noise, XGBoost gives slightly better results than LightGBM at higher noise levels.

Table 2. *Cheating Detection Recall under Different Noise Levels*

Noise Level	XGBoost	LightGBM	Random Forest
10%	0.98	0.98	1.00
20%	0.96	0.94	1.00
30%	0.90	0.87	1.00
40%	0.78	0.77	0.98
50%	0.51	0.51	0.53

At 50% noise, all models show a drop in performance (accuracy < 0.5) but Random Forest only achieves 0.47. Although Random Forest reached 53% recall. This shows that the data is too noisy to learn meaningful patterns and is almost at random prediction level. 50% error rate in labeling is too high for any practical application. We should focus on data quality instead of evaluating models.

5. Conclusion

This research evaluated three machine learning models for detecting cheating in online learning under different label noise levels. The results show that the ensemble method, specifically Random Forest, performs better than gradient boosting models in both clean and noisy data. This model also achieved the highest recall for detecting cheating and stayed strong even with 40% noise. Gradient boosting models also show the good result under low noise. If we can focus on data quality and the better labeling rules, XGBoost is a considerable option. However, our limitations include the small dataset size and subjective labeling based solely on our judgment, not directly observed. There may be caused false positives, such as genuinely fast/knowledgeable students misclassified. Our future research will take a closer look at the interpretability of these models to help educators understand the factors that contribute to cheating detection decisions. Besides, we will try to apply for other Machine

Learning models. A lager comparison will bring more details and more accurate conclusion about the best model for cheating detection.

References

Du, J., Song, Y., An, M., An, M., Bogart, C., & Sakr, M. (2022). Cheating Detection in Online Assessments via Timeline Analysis. Proceedings of the 53rd ACM Technical Symposium on Computer Science Education, 98–104. https://doi.org/10.1145/3478431.3499368

Erdem, B., & Karabatak, M. (2025). Cheating detection in online exams using deep learning and machine learning. Applied Sciences, 15(1), 400.

Garcines, K. M. A., Estender, M. R., Epondol, J. M., Uy, S. D., MaldepeÃ±a, K. M. V., & Cuevas, J. F. (2024). Stories behind Academic Cheating: Cheatersâ€TM Perspective. International Journal of Research and Innovation in Social Science, 8(12), 2013–2024.

Garg, M., & Goel, A. (2025). Towards Fair Assessments: A Machine Learning-based Approach for Detecting Cheating in Online Assessments. Proceedings of the 15th International Learning Analytics and Knowledge Conference, 104–114. https://doi.org/10.1145/3706468.3706482

Hassan Hosny, H. A., Ibrahim, A. A., Elmesalawy, M. M., & Abd El-Haleem, A. M. (2022). An intelligent approach for fair assessment of online laboratory examinations in laboratory learning systems based on Student's mouse interaction behavior. Applied Sciences, 12(22), 11416.

Jalilzadeh, K., Rashtchi, M., & Mirzapour, F. (2024). Cheating in online assessment: A qualitative study on reasons and coping strategies focusing on EFL teachers' perceptions. Language Testing in Asia, 14(1), 29. https://doi.org/10.1186/s40468-024-00304-1

Kamalov, F., Sulieman, H., & Santandreu Calonge, D. (2021). Machine learning based approach to exam cheating detection. Plos One, 16(8), e0254340.

McCabe, D. L., & Trevino, L. K. (1993). Academic Dishonesty: Honor Codes and other Contextual Influences. The Journal of Higher Education, 64(5), 522–538. https://doi.org/10.1080/00221546.1993.11778446

Sozon, M., Alkharabsheh, O. H. M., Fong, P. W., & Chuan, S. B. (2024). Cheating and plagiarism in higher education institutions (HEIs): A literature review. F1000Research, 13, 788.

Watson, G. R., & Sottile, J. (2010). Cheating in the digital age: Do students cheat more in online courses? https://mds.marshall.edu/eft_faculty/1/

Xiao, R., Huerta-Mercado, E., & Garcia, D. (2022). Detecting Cheating in Online Take-Home Exams with Randomized Questions. Proceedings of the 54th ACM Technical Symposium on Computer Science Education V. 2, 1323–1323. https://doi.org/10.1145/3545947.3576270

McDonnell, M., & Tantong, K. (2023). THE EVOLUTION OF ACADEMIC DISHONESTY: A REVIEW OF THE LASTEST TRENDS AND SOLUTIONS TO STUDENT CHEATING PRACTICES IN ONLINE EDUCATION. Journal of Education and Innovation, 25(3), 351–361.