Journal of Data Science and Modeling, Vol. 3, No. 1, 175-199, December 2024

Research Manuscript

# CheatingRank: A Multi-Stage Approach for Detecting Cheating in Online Assessments

Seyede Fatemeh Noorani *[1], Mohammad Reza Mohammadi[1]
Maryam Karimi[2], Nasrin Taherkhani[1]

[1]Department of Information and Computer Engineering, Faculty of Technology and Engineering, Payame Noor University (PNU), P.O. Box 19395-4697, Tehran, Iran

[2]Department of Computer Science, Faculty of Mathematical Sciences, Shahrekord University, Shahrekord 88186-34141, Iran

Recieved: 25/06/2025 Accepted: 01/03/2026

**Abstract:** With the rapid expansion of virtual education, detecting cheating in online examinations has emerged as a critical challenge in ensuring the validity and integrity of educational assessments. In this study, the concept of "Cheating Rank"(CR) is introduced, and this rank is calculated using a multi-stage method based on the analysis of data derived from online examinations. In this approach, students' answer sheets are optimally categorized using the $K-means++$ clustering algorithm, with the number of clusters dynamically and optimally determined, thereby distinguishing similar behavioral and performance patterns. Subsequently, by examining network status, examination duration, and response patterns, the cheating rank of each student is determined. To evaluate the validity of this ranking, the significance of the relationship between the performance discrepancy of students in in-person examinations and online examinations and the cheating rank is assessed using the non-parametric Mann-Whitney U test. The results demonstrate the high effectiveness and generalizability of the combined approach of dynamic clustering and multi-stage analysis in monitoring the integrity of online examinations and identifying cheating.

**Keywords:** Online Exam, Cheating Rank, Clustering, Fraud Detection.
**Mathematics Subject Classification (2010):** 99X99, 99X99.

*Corresponding Author: sf.noorani@pnu.ac.ir

# 1. Introduction

Over the past decade, the proliferation of digital technologies has engendered new opportunities in education. One of the most significant developments is the ability of students to enroll in online courses, thereby offering them a choice between traditional in-person classes and computer-based instruction. These online courses have rapidly gained popularity among students due to their enhanced flexibility and reduced commuting costs. In parallel, online examinations have emerged as a pivotal tool for assessing student performance, progress, and the quality of education within this digital framework. Such examinations not only evaluate educational objectives but also facilitate the learning process by providing prompt feedback Kaddoura and Gumaei (2022); Yulita and *et al.* (2023); Carrasco and *et al.* (2023); Noorbehbahani *et al.* (2022).

Online examinations provide valuable insights into the effectiveness of educational courses and play a pivotal role in identifying strengths and weaknesses in instructional design and delivery. Among the advantages of this assessment modality are its environmental sustainability, cost-efficiency, logistical convenience, rapid feedback provision, enhanced assessment accuracy, and improved educational decision-making. However, despite these notable benefits, online examinations are confronted with significant challenges, including an elevated risk of cheating and breaches of academic integrity Yulita and *et al.* (2023).

Such challenges can directly undermine the quality and credibility of online education, thereby eroding trust in this mode of instruction Ramzan and *et al.* (2024). Furthermore, the complexity of monitoring online examinations and the difficulty in controlling suspicious behaviors necessitate the adoption of technologically advanced methods for more precise assessment of test-takers performance. The integration of cutting-edge technologies with behavioral analysis algorithms can facilitate the detection of anomalous activities during examinations, contributing to the establishment of a fair and secure testing environment Wang and *et al.* (2023).

Weaknesses in internal control mechanisms and the trivialization of the consequences of cheating may create an environment conducive to increased cheating across various subjects and educational settings Chen and *et al.* (2024). Moreover, the issue of cheating is not confined to educational contexts alone. Extensive research has been conducted in domains such as financial fraud detection and the identification of fake social media accounts, where algorithms for fraud detection have been developed Lai and *et al.* (2023).

In the realm of education, numerous studies have explored strategies for cheating detection. For instance, one study Yulita and *et al.* (2023) proposed a model based on the analysis of students' physical positioning and movement patterns in front of a camera, coupled with neural network techniques, to predict the likelihood of cheating. Another study Borzoei and *et al.* (2021) employed deep learning models to analyze visual frames and speech data to identify fraudulent behaviors. Additionally, research Kaddoura and Gumaei (2022) investigated the relationship between the timing of exam initiation and students' multiple-choice responses to assess the probability of cheating among them.

In Essay Exams, conventional cheating detection methods are often insufficient, necessitating advanced text analysis techniques to identify linguistic patterns and examine inter-response correlations. For example, a study Cao and *et al.* (2023) utilized a method called $BTextCan$ for cheating detection, which leverages the $BERT\_BASE$ language model to map user responses into vectors and subsequently selects the most salient features from the extracted local attributes. While $BERT\_BASE$ is effective for processing large datasets, its efficiency diminishes in online written exams due to limited data availability and its focus on individual sentences, hindering its ability to capture deep contextual relationships.

In another study Wang and *et al.* (2015), a combination of decision tree algorithms and Gradient Boosting Machines (GBM) was employed for cheating detection. Despite its high accuracy, this approach faces challenges such as overfitting, extensive tree structures, and high sensitivity to noise, which adversely affect result quality.

Current machine learning methods for cheating detection predominantly rely on data sampling. Although this approach enables algorithms to observe diverse behaviors, random sample selection may yield unstable results and fail to provide a comprehensive understanding of behavioral patterns Xu and *et al.* (2023). Consequently, there is a pressing need for more precise and robust methodologies.

In current research, a novel method for detecting cheating in online examinations is proposed. Unlike existing methods that often require controlled environments or specialized hardware such as cameras, microphones, and high-speed internet, the proposed approach is based on a comprehensive analysis of exams logs data, including scores, response times, answer patterns, and network-related information. This enables cheating detection without the need for additional equipment. To achieve this, a multi-stage methodology has been developed, wherein the $CheatingRank$ (**CR**) of each student is quantitatively calculated based on behavioral and performance indicators.

The remainder of this article is structured as follows: Section 2 reviews studies related to cheating detection across various domains. Section 3 elaborates on the stages of the proposed cheating detection method. Section 4 evaluates the performance of the presented model and reports the results. Finally, Section 6 concludes the article with a summary of the research findings.

## 2. Literature Review

In recent years, fraud and anomaly detection have been extensively studied across various domains, including finance, telecommunications, digital currencies, and online platforms. In these domains, a wide range of techniques such as classification models, clustering algorithms, statistical analysis, explainable artificial intelligence, and deep learning architectures have been employed to identify suspicious patterns and abnormal behaviors. These methodological advances provide a solid foundation for analyzing large-scale and complex behavioral data.

Several studies have demonstrated the effectiveness of advanced machine learning and neural network models for fraud detection. For instance, attention-based and boosting-based models have been proposed to extract latent fraud patterns and improve detection performance Xu and *et al.* (2023). In addition, classification-based frameworks combining secure and dynamic clustering mechanisms have shown strong robustness in online environments Zhou and *et al.* (2023). Recent research has also explored explainable artificial intelligence models to enhance interpretability in financial fraud detection Yi and *et al.* (2023); Kazemi and Piri (2023); Mohammadi and *et al.* (2023), as well as supervised learning approaches for fraud detection in digital currency platforms Lei and *et al.* (2023). Moreover, advanced neural network architectures, including distributed neural networks and graph neural networks, have been successfully applied to detect fraud in complex and sparse datasets Chen and *et al.* (2024); Habibpour and *et al.* (2023); Ghamari Moghaddam and Nakhaei (2023). Hybrid and graph-based approaches have also been investigated in specialized domains such as tax fraud, online auctions, telecommunications, and network anomaly detection Dadfarnia and Adibnia (2019); Wu and *et al.* (2024); Jalali Dehkordi and *et al.* (2023); Perumal and *et al.* (2024). Furthermore, object detection-based monitoring systems combined with blockchain technologies have been proposed to reduce human supervision and prevent manipulation in exam environments Kulkarni and *et al.* (2025).

Building upon these methodological developments, cheating detection in educational assessments has attracted increasing attention. Early approaches focused on paper-based examinations, demonstrating that cheating behavior is influenced not only by answer similarity but also by the exam environment Hung and *et al.* (2010). With the transition to online and computer-based exams, technology-driven solutions such as personalized exam sheets and monitoring software have been introduced to mitigate cheating. Given the challenges of large-scale online monitoring, deep learning-based approaches using human activity recognition and webcam data have been proposed to automatically detect suspicious behaviors, achieving promising performance Yulita and *et al.* (2023).

In the context of online examinations and higher education, behavioral and response-based analyses have been shown to be particularly effective. Multi-layered feature analysis and response similarity metrics have demonstrated high accuracy in detecting cheating in online exams Duhaim and *et al.* (2021). Clustering techniques applied to features such as IP addresses, response times, and answer patterns have been successfully utilized in learning management systems such as Moodle to identify anomalous student behaviors Noroozi and Mohamady (2011). More recently, cheating detection has been explicitly investigated in higher education environments through the analysis of student activity and interaction data, highlighting the importance of behavioral indicators and scalable automated detection mechanisms Alsabhan (2023); Ortin and *et al.* (2025). In addition, real-time or near real-time cheating detection systems for online exams have been proposed to identify suspicious behaviors during assessments Khabbachi and *et al.* (2024).

Despite the diversity of existing methods, many approaches rely on intrusive real-time monitoring or single-stage detection strategies, which may limit their applicability across different courses and examination settings. Industry-oriented reports have also summarized common cheating strategies in online assessments, further emphasizing practical challenges faced by educational institutions Janison (2024). In contrast, this study introduces a multi-stage, education-oriented framework that employs dynamic $K$-means++ clustering to analyze online examination data, including student responses, response times, and network information. By assigning a quantitative Cheating Rank to each student, the proposed approach enables more precise behavioral analysis while enhancing robustness and generalizability.

Table 1: Comparison of Previous Studies

| Study | Method | Classification | Clustering | Mathematical | Other | Context | Evaluation Criteria |
|---|---|---|---|---|---|---|---|
| Yulita and *et al.* (2023) | HAR | ✓ | | | | Education | F1 |
| Wang and *et al.* (2023) | $SCN\_GNN$ | ✓ | | | | Financial | – |
| Cao and *et al.* (2023) | $AM\_Boost$ | ✓ | | | | Financial | Re,F1,AUC |
| Wang and *et al.* (2015) | $Monitoring$ | | | | ✓ | Education | – |
| Hung and *et al.* (2010) | $DCA, CSPRA$ | | | ✓ | | Education | – |
| Zhou and *et al.* (2023) | $XAI$ | ✓ | | | | Financial | – |
| Yi and *et al.* (2023) | $ESOA$ | ✓ | | | | Financial | $Precision, AUC$ |
| Kazemi and Piri (2023) | $SVM, DT$ | ✓ | | | | Financial | $Precision$ |
| Lei and *et al.* (2023) | $DDNN$ | ✓ | | | | Financial | $Precision$ |
| Habibpour and *et al.* (2023) | $Ensemble, MCD$ | ✓ | | | | Financial | $Precision$ |
| Ghamari Moghaddam and Nakhaei (2023) | $ID3$ | ✓ | | | | Tax | $Accuracy$ |
| Dadfarnia and Adibnia (2019) | $L1\_Noram$ | ✓ | | | | Auction | $TPR, FPR$ |
| Wu and *et al.* (2024) | $LSG\_FD$ | ✓ | | | | Network | |
| Jalali Dehkordi and *et al.* (2023) | $DMDC$ | ✓ | | | | Education | $F1, Preciaion$ |
| Duhaim and *et al.* (2021) | | | ✓ | | | Education | $Jaccard$ |
| Kulkarni and *et al.* (2025) | $PoseNet, COCO-SSD$ | ✓ | | | | Education | |
| Perumal and *et al.* (2024) | $CAEP$ | | ✓ | | | Network | $F1, precision$ |
| Alsabhan (2023) | $LSTM$ | ✓ | | | | Education | $Accuracy$ |
| Ortin and *et al.* (2025) | $RNN, CNN$ | ✓ | | | | Education | $Accuracy, F1$ |
| Khabbachi and *et al.* (2024) | $emotionrecognition$ | ✓ | | | | Education | $F1, Precision$ |
| Janison (2024) | $RNN, DFT$ | ✓ | | | | Education | $Accuracy, F1$ |
| $CurrentStudy$ | $MultiStage$ | | ✓ | | | Education | $Silhouette$ |

# 3. Methodology

In this research, the $CheatingRank(\mathbf{CR})$ is introduced, which is a numerical value ranging from 0 to 6, assigned based on students' online exam results and behavioral patterns. A Cheating Rank of 0 indicates that, according to the proposed method, no cheating has occurred. Higher values suggest a greater likelihood of the student exhibiting behaviors indicative of cheating.

To both prevent potential cheating and ensure that variations in internet speed do not affect the analysis, exam questions were presented to students one at a time and in a randomized order. This design serves as a routine safeguard while allowing fair evaluation of all students.

The online exam log file is first analyzed. The exam consists of multiple-choice questions. Initially, the $CR$ for all students is set to 0. Students are then clustered based on their responses to online exams. Within each cluster, if the similarity exceeds a predefined threshold, one unit is added to the corresponding student's $CR$.

In the next stage, the IP address used by the student to access the exam is examined. This analysis is conducted from two perspectives. First, the student's IP address is compared for similarity with those of other students. Second, the number of distinct IP addresses used by the student to access the same exam is evaluated. At this stage, up to 2 units may be added to the $CR$.

In the subsequent stage, the number of sessions a student has attended in instructional classes is evaluated. If this count falls below a predefined threshold, one unit is added to the $CR$.

Another factor examined is the time spent on the online exam. If a student completes the exam significantly faster than the allocated time, one unit is added to their $CR$.

The final stage involves analyzing the response behavior to exam questions. If a student answers questions correctly without altering any options, an additional unit is added to their $CR$.

Figure 1 illustrates the stages of the proposed method. Further details of this method are elaborated in the following sections. As with any methodological approach, evaluation is essential. To this end, an in-person exam with the same content as the online exam is administered to students at the end of the term. The model's performance is then assessed by comparing students' results in both the in-person and online exams and statistically correlating them with the derived $CR$.

Each of these sections will be explained in detail below.

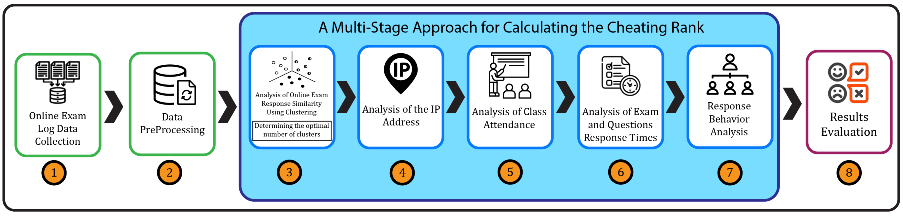



Figure 1: The Stages of Proposed Method.

## 3.1 Data Set

The dataset for this study comprises information from the Learning Management System (LMS) and online exam records of 140 undergraduate Computer Engineering students at Payame Noor University, Tehran Province. Among these students, 18 withdrew from the course before the final exams and were excluded from the analysis. The data of the remaining 122 students, along with the semester in which the exams were conducted, are presented in Table 2. As indicated in the table, three classes from two different courses were selected for analysis: "Discrete Mathematics" (one class) and "Automata Theory and Languages" (two classes in different semesters). These courses were delivered online via the university's LMS. Students completed one online exam and one in-person exam at the end of the semester. Both exams were designed using common course content, and each exam was scored out of 8 points.

Table 2: Details of the Three Classes Included in the Dataset

| CourseID | Course Name | Semester | Number of Students | Female | Male |
|---|---|---|---|---|---|
| *A* | Automata Theory and Languages | 4021 | 20 | 11 | 9 |
| *B* | Discrete Mathematics | 4021 | 45 | 19 | 26 |
| *C* | Discrete Mathematics | 4022 | 57 | 19 | 38 |
| | *Sum* | | 122 | | |

The online exam was configured such that students could view only one question at a time, with the next question appearing only after the current one was answered. To enhance the integrity of the online exam, questions were designed without a "back" option; once a response was submitted and the next question displayed, students could neither review nor revise previous answers.

Upon completion of the online exam, two log files are accessible. The first log file includes the following attributes:

- Username: Corresponds to the student ID.
- First Name

- Last Name
- Exam Taken: Indicates participation status (TRUE if the student participated, FALSE otherwise).
- Unanswered: Number of unanswered questions.
- Incorrect: Number of incorrectly answered questions.
- Correct: Number of correctly answered questions.
- Score: Final exam score.

Additionally, detailed response information for each individual exam question is also accessible in this file.

The second log file records activity history, including the time of exam entry, time of accessing each question, time of any option changes, and time of response submission. This log file comprises the following attributes:

- Date: Based on day and hour.
- Username: Student ID.
- IP: IP address of the device used by the student to access the exam.
- Log Type: Includes two values: "Activity Entry" and "Response Submission." The former indicates the time of entering the exam or a specific question, while the latter denotes the time of submitting a response.
- Details: Contains information about the course and the corresponding question.

Parts (a) and (b) of Figure 2 display excerpts from the first and second log files related to the exam, respectively.

Additionally, the number of online class sessions attended was extracted from the system, and this feature was also incorporated into the dataset generation process. By integrating the "number of online class sessions attended" and selecting specific attributes from the two online exam log files, the final set of features listed in Table 3 were employed for analysis in this study.

Ultimately, the data set comprised 1536 records, with the number of records corresponding to each course and semester presented in Table 4.

| نمره | صحیح | غلط | سفید | آزمون داده | نام خانوادگی | نام | نام کاربری |
|---|---|---|---|---|---|---|---|
| 4.50 | 3 | 2 | 0 | True | ابراهیمیان واقع نژاد | الهه | 402161023 |
| | | | | False | اسدي | حنانه | 402161058 |
| 4.50 | 3 | 2 | 0 | True | اصغریان سیدبیگلو | مهدیه سادات | 402161082 |
| 6.00 | 4 | 1 | 0 | True | اقبالي باشسیز | فهیمه | 400999378 |
| 4.50 | 3 | 2 | 0 | True | اکبرزاده کلاهی | ماهرخ | 400872879 |
| 4.50 | 3 | 2 | 0 | True | امیدي | فاطمه | 402167067 |
| 4.50 | 3 | 2 | 0 | True | ایزي | فاطمه | 990064799 |
| 1.50 | 1 | 4 | 0 | True | بازبندي | وحید | 963834156 |
| | | | | False | بخشی ایلانی | محدثه | 401874525 |
| 6.00 | 4 | 1 | 0 | True | براتي | حافظ | 401019582 |
| 6.00 | 4 | 1 | 0 | True | پوري اکرم | رکسانا | 401019596 |
| 4.50 | 3 | 2 | 0 | True | ترکاشوند | حانیه | 401037377 |
| 4.50 | 3 | 2 | 0 | True | ترکاشوند | حنانه | 401991206 |

(a) The first log file

| توضیحات | نوع لاگ | Ip | نام کاربری | تاریخ |
|---|---|---|---|---|
| match:116176-qid:57706-status:True | ثبت پاسخ | 5.112.61.219 | 401991400 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57701-status:True | ثبت پاسخ | 5.211.21.43 | 402214553 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 5.117.220.205 | 990161740 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57716-status:True | ثبت پاسخ | 93.118.122.253 | 401019596 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 188.211.190.120 | 402052909 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 5.113.230.146 | 400947630 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 5.112.242.73 | 402161082 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 5.117.220.205 | 990161740 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 5.117.220.205 | 990161740 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 188.211.190.120 | 402052909 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| آزمون میانترم | ورود به فعالیت | 46.41.233.75 | 400872879 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۹ |
| match:116176-qid:57713-status:True | ثبت پاسخ | 89.198.198.28 | 400875730 | چهارشنبه ۳۱ اردیبهشت ۱۴۰۴ - ۱۰:۳۸ |

(b) The second log file

Figure 2: The online exams log files

## 3.2 Data Preprocessing

Due to the variability in the values of different features within the dataset, normalization was necessary, specifically for time and IP address data.

For time, since the activity timestamps in the exam log are recorded in the format "Sunday, May 10, 2024 17:22," the hour and minute were first extracted. Subsequently, time was segmented into 5-minute intervals. For example, if the exam started at 17:00 and the student entered within the first 5 minutes, it was categorized as Interval 1. If they entered during the second 5-minute window (e.g., 17:07), it was categorized as Interval 2. This process was applied to the start time, response submission time, answer change time, and exam completion time.

For IP addresses, each IP value was converted into a decimal number. This was achieved by multiplying each segment of the IP address (from right to left) by $256^0, 256^1, 256^2, and 256^3$ respectively, and then summing the results. For example,

Table 3: The attributes of Dataset

| $No$ | $AttributeName$ | $Description$ |
|---|---|---|
| 1 | $StudentId$ | Student number |
| 2 | $ExamName$ | $(A, B, C)$ |
| 3 | Activity Type | Including the values "Exam Entry","Question Answered", "Answer Changed", and "Exam Completion". |
| 4 | Status of Questions 1 to 10 | 0:Unanswered or Incorrectly Answered<br>1: Correctly Answered<br>Each question is treated as a distinct attribute, resulting in a total of 10 attributes. |
| 5 | Number of Unanswered Questions | |
| 6 | Correct Answers Count | |
| 7 | Incorrect Answers Count | |
| 8 | Score | |
| 8 | Exam Start Time | |
| 9 | IP(S) | IPs Used by the Student to Access the Exam |
| 10 | Attendance Count | Number of Online Sessions Attended for the Course |
| 11 | Total Time Spent on Exam | |
| 12 | Answer Change Count per Question | The log file captures the count of answer option changes for each question. |

Table 4: Number of Records in the Dataset

| Course | Number of Students | Number of Records |
|---|---|---|
| A | 20 | 309 |
| B | 45 | 675 |
| C | 57 | 556 |

for the address 213.195.1.166:

$$213 \times 256^3 + 195 \times 256^2 + 1 \times 256^1 + 166 \times 256^0 = 3586326950$$

This normalization ensures consistency and facilitates further analysis.

### 3.3 A Multi-Stage Approach for Calculating the Cheating Rank

The proposed Cheating Rank calculation approach in this study consists of five distinct stages. In the first stage, response similarities in the online exam are examined. The second stage involves analyzing the IP addresses used by the student to access the exam. In the third stage, the student's participation and attendance in online classes for the relevant course are evaluated. The fourth stage assesses the speed of response during the exam, while the fifth stage analyzes the response approach, specifically focusing on whether answer options were changed during the exam.

At each stage, if suspicious indicators are detected, one unit is added to the student's $CR$. The $CR$ ranges from 0 to 6, where 0 indicates no cheating, and higher values signify an increased likelihood of cheating and more suspicious conditions.

In this research, initial clustering is performed solely based on responses to each online exam question, the number of correct and incorrect answers, and the exam score. This selection allows for better focus on direct indicators of cheating and enables a gradual, step-by-step analysis.

Thus, this multi-stage approach ensures a balance between model accuracy and efficiency while providing deeper analytical insights and improved interpretability of results.

The details of these stages are elaborated in the following sections.

### 3.3.1 Analysis of Online Exam Response Similarity Using Clustering

In this stage, clustering is performed using the following features: correct or incorrect responses to each multiple-choice question, number of unanswered questions, number of incorrect answers, number of correct answers, actual exam start time for each student, and score. While the official exam start time is the same for all students, the recorded start time reflects "when each student actually entered the exam", which can vary among students. For each question, a value of 1 indicates that the student answered correctly, while 0 indicates either an unanswered or incorrectly answered question. The K-means++ algorithm is employed for clustering based on these features.

#### 3.3.1.1 K-means++ algorithm:

The K-means++ algorithm is an intelligent enhancement of the traditional K-means algorithm, designed to improve the selection of initial cluster centers. In standard K-means, initial centers are chosen randomly, which can lead to poor convergence or suboptimal results. In contrast, K-means++ selects initial centers in a way that optimizes their distribution and ensures they are well-separated. This smart initialization results in faster convergence and higher-quality clustering compared to traditional K-means. The general process of the K-means++ algorithm is as follows:

- First Center Selection: One center is randomly selected from the dataset.

- Subsequent Center Selection: Each subsequent center is chosen with a probability proportional to the squared distance from the nearest already-selected center. This means data points farther from existing centers have a higher probability of being selected as the next center.

- Repeat Until k Centers Are Selected: This process continues until all k centers are chosen.

By ensuring a more balanced and strategic placement of initial centers, K-means++ significantly improves the clustering outcome.

### 3.3.2 Stage 1: Cheating Rank Assignment within Clusters

To select clusters for analyzing in subsequent stages, the average similarity between students within each cluster is calculated. Subsequently, leveraging expert opinion (from the course instructor) and the Pareto Principle, the likelihood of cheating within clusters is assessed.

The Pareto Principle, also known as the 80/20 rule, was first introduced by Italian economist Vilfredo Pareto. It posits that in many systems and natural or social phenomena, outcomes are distributed asymmetrically, with a small proportion of factors accounting for a large proportion of results. Specifically, the principle suggests that approximately 20% of causes are responsible for 80% of effects. While the exact 80/20 ratio may not hold in every context, the core concept—namely, the asymmetry in cause-and-effect distributions—is observable across domains such as economics, management, data analysis, and education. In the context of cheating detection in online exams, this principle can be applied to identify a small subset of participants who exhibit the highest deviation or likelihood of cheating relative to others. Thus, incorporating the Pareto Principle in the design of cheating detection systems can help focus analytical resources on higher-risk cases, thereby enhancing detection efficiency Newbold and *et al.* (2013).

Based on this approach, the distance between each cluster member and the other members is calculated. If the average similarity of online exam responses within a cluster exceeds 80%, that cluster is deemed suspicious of cheating, and the $CR$ of 1 is assigned to all its members.

### 3.3.3 Stage 2: Analysis of the IP Address Used by the Student to Access the Exam

In this study, students sharing the same IP address were distinguished from other participants. It is important to note that shared IP addresses do not necessarily imply fraudulent behavior; for example, students accessing the exam from a common location, such as dormitories via the same network, may do so without any misconduct. However, it cannot be overlooked that multiple students taking an exam in an unsupervised shared environment increases the risk of cheating Gao (2012).

Additionally, any student using more than one IP address during the exam was considered at risk of misconduct. The use of multiple IPs by a single individual during the exam may indicate non-transparent behavior inconsistent with exam regulations. Consequently, this is defined as a violation of exam rules and contributes to an increase in the Cheating Rank Tiong and *et al.* (2022).

Accordingly, up to 2 units may be added to the $CR$ in this stage. If students accessed the exam using a shared IP, 1 unit is added to their $CR$. Subsequently, any student who accessed the exam using multiple IPs also receives an additional 1 unit to their $CR$.

### 3.3.4 Stage 3: Analysis of Class Attendance

In educational settings, assessing student behavior and performance is crucial for evaluating learning outcomes. Frequent absences or insufficient class participation can serve as indicators of increased cheating likelihood Valızadeh (2022). To account for the impact of class attendance on the Cheating Rank, a threshold is defined; students with attendance below this threshold receive an additional unit to their $CR$. This threshold is determined using both the instructor's expertise and the Pareto Principle.

According to the Pareto Principle, approximately 20% of students account for the majority of irregularities. Therefore, in this study, the attendance threshold is set at 80%, meaning students attending fewer than 80% of sessions are considered at higher risk of cheating, and their $CR$ is increased by one unit.

### 3.3.5 Stage 4: Analysis of Exam and Questions Response Time

In this study, the time spent by students on answering the total exam questions was examined as a behavioral indicator. If a student's total response time is significantly less than the standard time set by the instructor, this situation may be considered a sign of potential cheating Tiong and *et al.* (2022). However, it is noteworthy that reduced time can sometimes be attributed to the student's proficiency. Therefore, to avoid misinterpretations, only when this reduction exceeds 20%, one unit is added to the Cheating Rank. This threshold was determined using the instructor's expertise and the Pareto Principle, with the value set at 80% of the total standard exam time.

In addition to the total exam duration, the time spent on each question was also analyzed. To this end, a standard method in data mining, namely outlier detection, was employed. In this method, a coefficient called the Interquartile Range (IQR) is used, which is defined as the difference between the first and third quartiles of the data. The first quartile (Q1) represents the point below which 25% of the data fall. The second quartile (Q2) corresponds to the median, where 50% of the data are located, and the third quartile (Q3) represents the point below which 75% of the data fall.

According to this approach, observations that are less than $Q1 - 1.5 \times IQR$ or greater than $Q3 + 1.5 \times IQR$ are considered outliers Han *et al.* (2012).

Accordingly, the first, second, and third quartiles of the time spent on each question are computed. Then, for each student, the time spent on each question is examined to determine whether it is an outlier.

Finally, if more than two questions for a given student are identified as outliers, one unit is added to that student's CR.

#### 3.3.6 Stage 5: Response Behavior Analysis

This study focuses on analyzing students' response patterns to identify suspicious behaviors and detect potential cheating in online exams. Since fraudulent behavior often exhibits patterns distinct from honest behavior, data analysis is employed to extract behavioral indicators. One such indicator is the pattern of answer changes. Students who hesitate and change their answers are considered less likely to be cheating. Conversely, as noted by Man (2023) Man and Harring (2023), rapid, unchanged responses with high scores may indicate abnormal behavior.

Importantly, this factor is interpreted in the context of overall student behavior and relative patterns within the class, taking into account differences in ability and exam difficulty. Notably, the exam questions in this study were of moderate difficulty, so correct answers without changes are not automatically suspicious. Accordingly, students who answered more than 80% of questions quickly and without changing responses were assigned an additional unit in their $CR$, reflecting a higher likelihood of irregular behavior, but not a definitive conclusion of cheating.

#### 3.3.7 Simulation Environment

In this study, Python programming language and Visual Studio Code (VSCode) were used for data preprocessing, clustering, and Cheating Rank calculation. VSCode is a versatile integrated development environment (IDE) with numerous extensions supporting project execution and simulation.

## 4. Results and Discussion

In this section, we introduce the performance metrics, describe the model parameter determination process, and compare the performance results of the proposed model with existing methods.

## 4.1 Clustering Performance Metrics

One important metric for evaluating clustering quality is the Silhouette Index Rousseeuw (1987), which quantitatively measures the compactness and separation of clusters. This metric assesses the quality of clustering by evaluating how close each sample is to its own cluster center relative to the nearest neighboring cluster. For each data sample, the silhouette value is calculated based on two parameters:

- a(i): The average distance of sample i to all other samples within its cluster (compactness metric).
- b(i): The average distance of sample i to all samples in the nearest neighboring cluster (separation metric).

The silhouette value for sample i is computed as follows:

$$S(i) = \frac{b(i) - a(i)}{max(a(i), b(i))} \tag{4.1}$$

The value of s(i) ranges from -1 to 1. A value closer to 1 indicates that the sample is well-clustered, suggesting high-quality clustering. A silhouette value near 0 implies that the sample lies between two clusters. If the value is negative, it suggests that the sample may have been misclassified. The average silhouette value across all samples serves as a comprehensive metric for evaluating clustering quality, with higher values indicating better clustering performance.

### 4.1.1 Model Parameter Determination

In this study, the number of clusters $(K)$ is determined dynamically and adaptively using the Elbow Method. This method involves running the $K - means++$ algorithm for various values of $K$ and calculating the Sum of Squared Errors $(SSE)$. By plotting the relationship between $K$ and $SSE$, the optimal $K$ is identified at the point where the rate of error reduction noticeably decreases (the "elbow point" or inflection point in the graph). The minimum value of K is set to 2, categorizing students into two groups: cheating and non-cheating. Conversely, the maximum $K$ occurs when each cluster contains only two students who have cheated together.

In research Duhaim and *et al.* (2021), standard K-means clustering was used to categorize student responses into fixed clusters to reduce cheating in online exams. In contrast, this study employs the Elbow Method to dynamically determine the optimal number of clusters for each exam based on data-driven criteria. To achieve this, $SSE$ values were computed for $K$ in the range $2 \leq k < \frac{Sl}{2}$, where $Sl$

represents the number of students. Figure 3 illustrates the clustering process results, showing varying $K$ values and corresponding $SSE$ values for three distinct courses. Based on Elbow Method analysis, the elbow points, indicating optimal cluster counts, were identified as $k = 5$ for Classes $A$ and $B$, and $k = 8$ for Course $C$. These points reflect a balanced trade-off between model complexity and clustering accuracy.

#### 4.1.2 Performance Comparison of the Proposed Method

In contrast to study Duhaim and *et al.* (2021), which employed the standard $K-means$ algorithm, this paper utilizes the $K-means++$ algorithm. Figure 4 compares the performance of $K-means$ and $K-means++$ in clustering student exam responses.

The horizontal axis represents varying $k$ values, while the vertical axis shows the silhouette score. As illustrated in Figure 4, the silhouette index values for $K-means++$ (red dashed lines) are consistently higher than those of $K-means$. This demonstrates the superiority of the proposed method in the initial clustering stage.

The clustering results for each selected course are presented in Table 5. As shown in the table, for example, in Course A, 5 clusters were formed, with the number of students in each cluster specified.

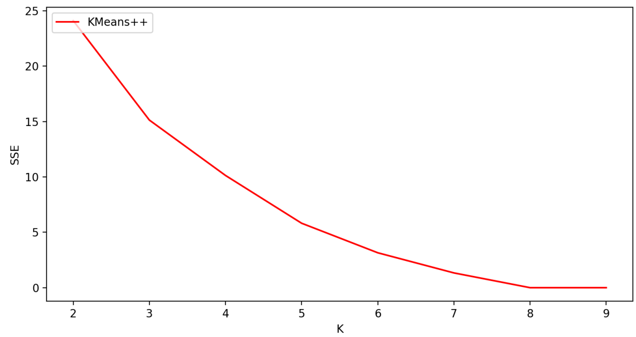
After clustering student responses, as described in the methodology section, the Cheating Rank is determined by analyzing IP addresses, educational participation, exam timing, and response behavior.

The results of this supplementary analysis are presented in Table 6.

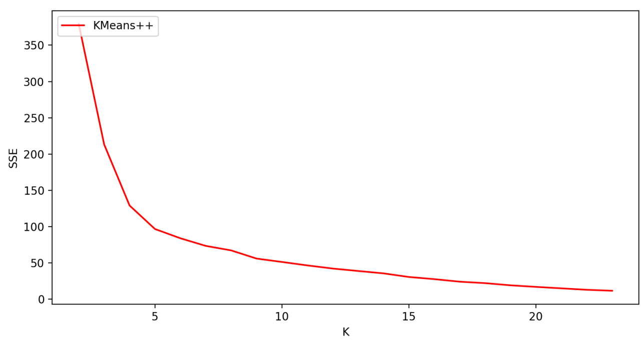
The Cheating Rank ranges from 0 to 6, where 0 indicates no cheating, and 6 signifies that all cheating conditions defined by the research method have been met. One student from Course A, 8 students from Course B, and 7 students from Course C received a Cheating Rank of 0, indicating no cheating according to the proposed method. Additionally, the results show that only one student was assigned a Cheating Rank of 4, and no students were identified with a Cheating Rank of 5 or 6.

Additionally, Figure 5 presents the aggregated results of online exam data analysis across all courses. According to the figure, a total of 16 students were assigned a Cheating Rank of 0 across the three courses, indicating no cheating based on the analysis. Seven students received a Cheating Rank of 1, 62 students a rank of 2, 36 students a rank of 3, and only one student was identified with a rank of 4. Based on the analysis conducted in this study, no students were detected with a Cheating Rank of 5 or 6 in the dataset.

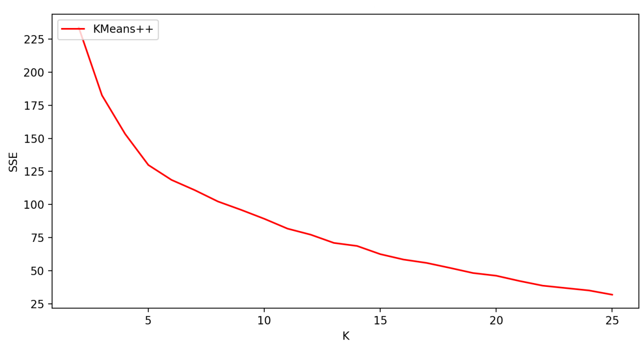
To evaluate the clustering results and validate the $CR$ derived from the pro-




(a) Class A

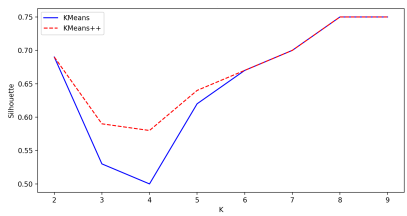



(b) Class B

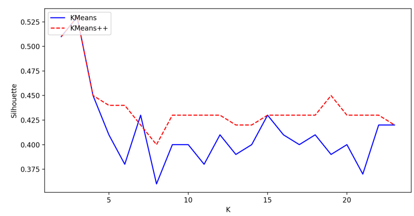



(c) Class C

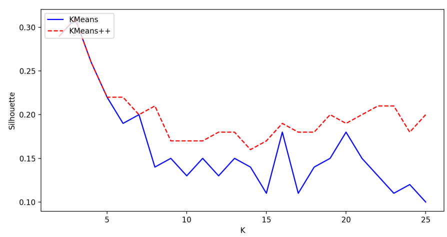
Figure 3: Determining the Optimal K Using the Elbow Method




(a) Class A




(b) Class B




(c) Class C

Figure 4: Comparison of K-means and K-means++ in the First Stage of the Proposed Method

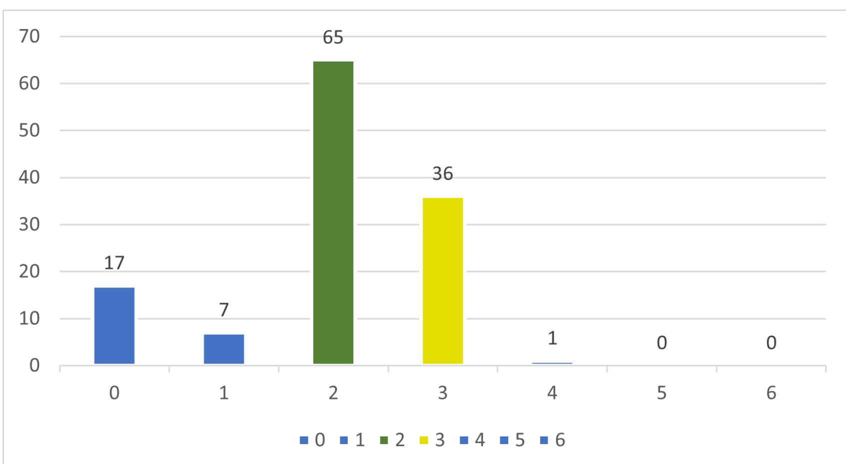



Figure 5: Number of Students Detected in Each Cheating Rank in This Study

posed method in this study, the absence of reliable labels in the dataset precludes the use of supervised classification evaluation methods. Therefore, in-person exam scores, conducted under controlled and cheat-free conditions, were employed as reference data for validation. This approach aligns with systematic clustering val-

Table 5: Clustering Details for Classes and Selected Clusters for Next Cheating Rank Stages

| Class | Clusters' Number | Number of Students in Cluster | Selection for Subsequent Stages |
|---|---|---|---|
| A | 0 | 4 | ✓ |
| | 1 | 1 | – |
| | 2 | 3 | ✓ |
| | 3 | 10 | ✓ |
| | 4 | 2 | ✓ |
| B | 0 | 6 | |
| | 1 | 14 | – |
| | 2 | 9 | ✓ |
| | 3 | 14 | ✓ |
| | 4 | 2 | |
| C | 0 | 10 | ✓ |
| | 1 | 23 | ✓ |
| | 2 | 3 | ✓ |
| | 3 | 7 | ✓ |
| | 4 | 4 | ✓ |
| | 5 | 1 | |
| | 6 | 3 | ✓ |
| | 7 | 6 | |

Table 6: Cheating Ranks Detected by the Proposed Method in This Study

| Course | Cheating Rank | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| A | 2 | 2 | 9 | 8 | 0 | 0 | 0 |
| B | 8 | 5 | 24 | 8 | 1 | 0 | 0 |
| C | 7 | 0 | 32 | 20 | 0 | 0 | 0 |
| Sum | 16 | 7 | 62 | 36 | 1 | 0 | 0 |

idation frameworks that leverage independent validation data, which have been established as a reliable standard for assessing unsupervised clustering results in recent studies Ullmann and *et al.* (2022).

Assuming the in-person exam scores as an objective and reliable criterion, students were categorized into two groups: "$SuspiciousofCheating$" and "$Normal$", based on both online and in-person exam scores. Both exams covered identical course content, with a maximum score of 8. The classification was based on two thresholds ($T\_online, T\_offline$): students whose online score exceeded $T\_online$ while their in-person score fell below $T\_offline$ were flagged as suspicious of cheating. To ensure that these thresholds reflect meaningful deviations rather than arbitrary cut-offs, $T\_online$ and $T\_offline$ were chosen based on the empirical score distributions, with online scores in the upper portion and in-person scores in the lower tail of the respective distributions.This approach is consistent with studies that use in-person exam scores as a valid benchmark for evaluating student performance and detecting anomalies in online exams, enabling group differentiation via score thresholds Dendir and Maxwell (2020).

In this study, two threshold scenarios were examined:

- Scenario 1: $T\_online = 5, T\_offline = 4$. Students with an online score > 5 and an in-person score < 4 were classified as suspicious.
- Scenario 2: $T\_online = 5, T\_offline = 3$. Here, students with an online score >5 and an in-person score <3 were flagged as suspicious, reflecting higher sensitivity to score discrepancies.

The Cheating Rank, calculated as a continuous variable from online exam behavioral data, served as the dependent variable for group comparison. Due to data skewness and unequal sample sizes between groups, the non-parametric $Mann - WhitneyU$ test was employed—a widely used method for comparing rank differences between two independent groups.

For Scenario 1, $(T\_online = 5, T\_offline = 4)$, 95 students were labeled as suspicious of cheating, and 27 as normal. The labeled dataset (suspicious vs. normal) was compared with the Cheating Rank using the $Mann - WhitneyU$ test. The analysis results are as follows:

Table 7: Results of the Mann-Whitney U Test for Exam Data with T_online=5 and T_offline=4

| Types | N | Mean Rank | Sum of Ranks |
|---|---|---|---|
| Suspicious | 95 | 65.40 | 6213 |
| Normal | 27 | 47.78 | 1290 |

The $Mann - WhitneyU$ test yielded a statistic of $U = 912.000, aZ - score = -2.475$, and a two-tailed $p - value = 0.013$. Given that $p < 0.05$, the difference in Cheating Ranks between the "Suspicious of Cheating" and "Normal" groups is statistically significant. Specifically, the "Suspicious of Cheating" group exhibits significantly higher Cheating Ranks compared to the "Normal" group.

For Scenario 2, $(T\_online = 5, T\_offline = 3)$,, students with online scores >5 and in-person scores <3 were labeled as suspicious. In this case, 63 students were flagged as suspicious, and 59 as normal. The labeled dataset was compared with the Cheating Rank using the $Mann - WhitneyU$ test. The analysis results are as follows:

The Mann-Whitney U test yielded a statistic of $U = 1537.500, Z - score = -1.782$, and a two-tailed $p - value = 0.075$. Although the $p - value$ is slightly

Table 8: Results of the Mann-Whitney U Test for Exam Data with T_online=5 and T_offline=3

| Types | N | Mean Rank | Sum of Ranks |
|---|---|---|---|
| Suspicious | 63 | 66.60 | 4195.50 |
| Normal | 59 | 56.06 | 3307.50 |

above 0.05, considering a significance level of $\alpha = 0.10$, the difference in mean ranks between the two groups approaches the threshold of significance. Thus, with 90% confidence, the Suspicious of Cheating group exhibits higher Cheating Ranks compared to the Normal group, indicating a trend toward significant difference.

In summary, the Mann-Whitney U test results demonstrate that the Cheating Rank, derived from online exam behavioral data, significantly differentiates between students suspicious of cheating and normal students. This validates the proposed method's effectiveness in identifying students likely to exhibit abnormal behavior in online exams. In other words, the Cheating Rank serves as a reliable quantitative indicator for detecting students suspected of cheating.

## 5. Conclusion

In this study, a multi-stage approach based on online exam data analysis was proposed to identify and rank students' Cheating Ranks (CR). Initially, student response sheets were clustered using the $K$-means++ algorithm with dynamic determination of $K$, forming groups with similar behavioral and performance characteristics. Subsequently, each student's Cheating Rank was determined based on various behavioral and performance parameters, including response patterns, network status, and exam duration.

The results demonstrated that the Cheating Rank effectively differentiates between students suspicious of cheating and normal students, with statistically significant differences between the two groups. These findings underscore the validity and efficacy of the proposed method in detecting cheating in online exams. Furthermore, the impact of threshold selection for group differentiation highlighted the importance of appropriate parameter tuning in enhancing detection accuracy and influencing model sensitivity and reliability. Using in-person exam scores as a reference criterion strengthened the credibility of the results and enabled more objective model evaluation.

Given the limitations in sample size and data scope, future research should focus on larger, more diverse, and multidimensional datasets to improve both diagnostic accuracy and model generalizability. Integrating the proposed approach with advanced machine learning techniques and behavioral analytics could significantly enhance online exam integrity monitoring systems.

Overall, this research represents a significant step toward developing cost-effective, reliable, and efficient methods to reduce cheating in online educational environments, fostering greater trust among instructors and students in assessment processes, and providing a foundation for future studies to integrate richer behavioral metrics for improved detection accuracy.

## Acknowledgment

We extend our gratitude to the students of Payame Noor University for their cooperation in this research.

## References

Alsabhan W. (2023), Student Cheating Detection in Higher Education, *MDPI Journal of Educational Research,* **15**, 1023–1035.

Arthur D., Vassilvitskii S. (2006), $k - means + +$ : *The advantages of careful seeding*, Stanford.

Borzoei M.R., Shahbazipoor M.M., and Mostofi Sharabiyani M.M. (2021), Opportunities and challenges of online education in the university education system, *Research in Islamic education issues,* **29**(51), 95-120.

Cameron A. C. and Trivedi P. K. (2013), *Regression Analysis of Count Data*, Second Edition, NewYork, Cambridge University Press.

Cao R., Wang J., Mao M., Liu G., Jiang C. (2021), Feature-wise attention based boosting ensemble method for fraud detection, *Engineering Applications of Artificial Intelligence,* **126**, 106975.

Carrasco M., Silva A. R., and Henriques R. (2023), Detecting fraudulent student communication in a multiple choice online test environment, *Journal of the American Statistical Association,* **81**, 338-346.

Chen J., Chen Q., Jiang F., Guo X., Sha K., and Wang Y. (2024), $SCN\_GNN$ : A GNN-based fraud detection algorithm combining strong node and graph topology information, *Expert Systems with Applications,* **237**, 121643.

Dadfarnia M., and Adibnia F.(2023), Collusive Fraud Classification in Network of Online Auction Using Similarity Measure in Collective Classification, *Electronic and Cyber Defense,* **7**(1), 95-103.

Dendir S., and Maxwell R.S. (2020), Cheating in online courses: Evidence from online proctoring, *Computers in Human Behavior Reports,* **2**, 100033.

Duhaim A. M., Al-Mamory S., and Mahdi M. S.(2024), Cheating detection in online exams during COVID-19 pandemic using data mining techniques, *Webology,* **19**(1), 341-366.

Feller W. (1972), *An Introduction to Probability Theory and Its Applications*, New York, John Wiley.

Gao Q.(2012), Using IP addresses as assisting tools to identify collusions, *International Journal of Business, Humanities and Technology,* **2**(1), 70-75.

Gelman A., Jakulin A., Grazia Pittau M. and Su Y. (2008), A weakly informative default prior distribution for logistic and other regression models, *Annals of Applied Statistics,* **4**, 1360-1383.

Ghamari Moghaddam A., and Nakhaei H. (2023), Providing a solution for predicting tax fraud companies based on optimized decision tree, support vector machine and Bayesian network, *Information technology innovations and applied Communications,* **2**(2), 15-32.

Habibpour M., Gharoun H., Mehdipour M., Tajally A., Asgharnezhad H., Shamsi A., Khosravi A., and Nahavandi S.(2023), Uncertainty-aware credit card fraud detection using deep learning, *Engineering Applications of Artificial Intelligence,* **123**, 106248.

Han J., Kamber M., and Pei J. (2012), *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, San Francisco.

Huang R., Tawfik H., Nagar A. (2010), A novel hybrid artificial immune inspired approach for online break-in fraud detection, *Procedia Computer Science,* **1**(1), 2733-2742.

Jalili Dehkordi F., Khosravi Farsani H., and Javdani Gandomani T. (2023), A New Approach to Cheating Detection in Online Exams [In Persian], *Computing and distributed systems,* **5**(2), 102-112.

Janison (2024), 6 Ways Students Cheat Online Assessments, and How to Prevent It, *Janison Resources,* Available online: https://www.janison.com/resources/post/6-common-ways-students-cheat-online-assessments (accessed on 22 January 2026).

Kaddoura S. and Gumaei A. (2022), Towards effective and efficient online exam systems using deep learning-based cheating detection approach, *Intelligent Systems with Applications,* **16**, 200153.

Kazemi T., Piri P. (2023), Financial Reporting Fraud Scheme Prediction via Machine Learning Approach–Multiclass Classification, *Empirical Research in Accounting,* **12**(4), 255-280.

Khabbachi I., Rahman M., and Fernandez P. (2024), A Cheating Detection System in Online Exams through Real-Time Analysis, *IEEE Access,* **12**, 11234–11247.

Kulkarni M.D., Awate A., Shahade M., and Nandwalkar B. (2025), ExamGuard: Smart contracts for secure online test, *Information Systems,* **128**, 102485.

Lai S., Wu J., Ma Z., Ye C. (2023), *BTextCAN* : Consumer fraud detection via group perception, *Information Processing & Management,* **60**(3), 103307.

Lei Y., Ma C., Ren Y., Chen X., Narayan S., and Huynh A. N. Q.,(2023), A distributed deep neural network model for credit card fraud detection, *Finance Research Letters,* **58**, 104547.

Man K., and Harring J. R.(2023), Detecting preknowledge cheating via innovative measures: A mixture hierarchical model for jointly modeling item responses, response times, and visual fixation counts, *Educational and Psychological Measurement,* **83**(5), 1059-1080.

Mohammadi S., Sharifi Mehr F., and Seydi Ghomshe M. (2023), Fraud Detection in Digital Currencies (Bitcoin) Using Data Mining and Fraud Prevention Strategies [In Persian], *The 9th International Conference on Management Sciences and Accounting,* **12**.

Newbold P., Carlson W.L.,Thorne B. M. (2013), *Statistics for business and economics*, Pearson.

Noorbehbahani F., Mohammadi A., and Aminazadeh M. (2022), A systematic review of research on cheating in online exams from 2010 to 2021, *Education and Information Technologies,* **27**, 8413–8460.

Noroozi Z., and Mohamady E., Detection and correction of cheat in the secret sharing schemes with ternary codes, *Journal of Passive Defence Science and Technology,* **1**(2), 5.

Ortin F., Smith J., and Lee H. (2025), Assistant for the Detection of Potential Cheating Behavior in Educational Settings, *Proceedings of Educational Data Mining*, **17**, 45–58.

Perumal S., Sujatha P. K., K. S., and Krishnan M. (2024), A deep unsupervised learning paradigm for network anomaly detection, *Journal of Network and Computer Applications,* **128**, 102485.

Ramzan M., Abid A. , Bilal M., Amir K. M., Memon S. A., and Chung T. (2024), Effectiveness of pre-trained CNN networks for detecting abnormal activities in online exams, *IEEE Access,* **12**, 21503-21519.

Ranger J., Schmidt N., Wolgast A. (2020), The Detection of Cheating on E-Exams in Higher Education: The Performance of Several Old and Some New Indicators, *Frontiers in Psychology,* **11**, Article 568825.

Rousseeuw, P. J. (1986), Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics,* **20**, 53–65

Tiong L.C.O, Lee H. J., and Lim K. L. (2022), Online assessment misconduct detection using internet protocol and behavioural classification, *arXiv preprint arXiv:2201.13226*.

Ullmann T., Hennig C., and Boulesteix A.L. (2022), Validation of cluster analysis results on validation data: A systematic framework, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* **12**(3), e1444.

Valizadeh M.(2022), CHEATING IN ONLINE LEARNING PROGRAMS: LEARNERS'PERCEPTIONS AND SOLUTIONS, *Turkish Online Journal of Distance Education,* **23**(1), 195-209.

Wang J., Tong Y., Ling M., Zhang A., Hao L., and Li X. (2015), Analysis on test cheating and its solutions based on extenics and information technology, *Procedia Computer Science,* **55**, 1009-1014.

Wang G., Ma J.,Chen G. (2023), Attentive statement fraud detection: Distinguishing multimodal financial data with fine-grained attention, *Decision Support Systems,* **167**, 113913.

Wolter, W. (1986), Some coverage error models for census data. *Journal of the American Statistical Association,* **81**, 338-346

Wu J., Hu R., Li D., Ren L., Huang Z., and Zang Y.(2024), An improved telecom fraud detection approach based on latent synergy graph learning, *Neural Networks,* **169**, 20-31.

Xu B., Wang Y., Liao X., and Wang K. (2023), Efficient fraud detection using deep boosting decision trees, *Decision Support Systems,* **175**, 114037.

Yi Z.,Cao X.,Pu X., Wu, Y., Chen Z., Khan A., Francis A.,Li, S. (2023), Fraud detection in capital markets: A novel machine learning approach, *Expert Systems with Applications,* **231**, 120760.

Yulita I. N., Hariz F. A., Suryana I. and Prabuwono A. S. (2023), Educational Innovation Faced with COVID-19: Deep learning for online exam cheating detection, *IEEE Transactions on Computational Social Systems,* **11**, 1108-1120.

Zhou Y., Li H., Xiao Z., and Qiu J. (2023), A user-centered explainable artificial intelligence approach for financial fraud detection, *Finance Research Letters,* **58**, 104309.