دوفصلنامه ایرانی آموزش از دور

سال هشتم، شماره اول، بهار و تابستان ۱۴۰۵ (۱۳-۲۶)

DOI: ۱۰,۳۰۴۷۳/IDEJ.۲۰۲٦,۷۷٦۹۹,۱۲۷٥

«مقاله پژوهشی»

کاربست الگوریتم XGBoost در تشخیص تکالیف مبتنی بر هوش مصنوعی: راهکاری نوین برای حفظ یکپارچگی آکادمیک در آموزش از راه دور

پیمان الماسی‌نژاد، محمد ذهابی

۱. گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور، تهران، ایران.

۲. گروه مهندسی کامپیوتر و فناوری اطلاعات، دانشکده فنی و مهندسی، دانشگاه پیام نور، تهران، ایران.

نویسنده مسئول:
پیمان الماسی‌نژاد
رایانامه: p.almasinejad@pnu.ac.ir

تاریخ دریافت: ۱۴۰۵/۰۱/۱۱
تاریخ پذیرش: ۱۴۰۵/۰۴/۱۹

چکیده

ظهور فناوری‌های نوین زبانی مبتنی بر هوش مصنوعی، اعتبارسنجی اصالت تکالیف در نظام‌های آموزش از راه دور را با دشواری جدی روبه‌رو ساخته است. در این پژوهش، راهکاری مبتنی بر الگوریتم XGBoost ارائه می‌شود که قادر است متون تولیدشده توسط هوش مصنوعی را از نوشته‌های انسانی بازشناسی کند. این روش با بهره‌گیری از ۲۴ ویژگی شامل دو گروه شاخص‌های سبک‌شناختی (واژگانی، نحوی و نگارشی) و شاخص‌های رفتاری (زمانی، ویرایشی و تعاملی) و با استفاده از مجموعه‌داده استاندارد HC۳ شامل ۱۰ هزار نمونه، مورد ارزیابی قرار گرفت. فرآیند اعتبارسنجی به روش ۵-تایی انجام شد و نتایج حاکی از عملکرد مطلوب مدل با نرخ درستی ۹۶/۱ درصد، دقت مثبت ۹۵/۸ درصد، حساسیت ۹۶/۴ درصد و نمره ترکیبی F1 برابر ۹۶/۱ درصد است. تحلیل اهمیت شاخص‌ها نشان داد که نسبت نوع به شمار واژگان (TTR) و میانگین درازای جمله، تأثیرگذارترین مؤلفه‌ها در فرآیند تشخیص هستند. چارچوب پیشنهادی می‌تواند ابزاری کارآمد برای مدرسان آموزش از راه دور در جهت پایش اصالت تکالیف و صیانت از انضباط علمی در دوران هوش مصنوعی تلقی گردد.

واژه‌های کلیدی

انضباط علمی، یادگیری الکترونیکی، XGBoost، سبک‌سنجی متن، ردگیری رفتار نوشتاری.

استناد به این مقاله:

الماسی‌نژاد، پیمان و ذهابی، محمد (۱۴۰۵). کاربست الگوریتم XGBoost در تشخیص تکالیف مبتنی بر هوش مصنوعی: راهکاری نوین برای حفظ یکپارچگی آکادمیک در آموزش از راه دور. دوفصلنامه ایرانی آموزش از دور، ۸(۳)، ۱۳-۲۶.
(DOI: ۱۰,۳۰۴۷۳/IDEJ.۲۰۲٦,۷۷٦۹۹,۱۲۷٥)

حق انشار این مستند، متعلق به نویسندگان آن است. ۱۴۰۵ ©. ناشر این مقاله، دانشگاه پیام نور است.

این مقاله تحت گواهی زیر منتشرشده و هر نوع استفاده غیرتجاری از آن مشروط بر استناد صحیح به مقاله و با رعایت شرایط مندرج در آدرس زیر مجاز است.

Creative Commons Attribution-NonCommercial ٤,٠ International license (https://creativecommons.org/licenses/by-nc/٤,٠/)

Iranian Distance Education Journal Open Access

Spring & Summer (٢٠٢٦) ٨(١): ١٣-٢٦
DOI: ١٠,٣٠٤٧٣/IDEJ.٢٠٢٦,٧٧٦٩٩,١٢٧٥

ORIGINAL ARTICLE

Application of the XGBoost Algorithm in Detecting AI-Generated Assignments: A Novel Approach to Preserving Academic Integrity in Distance Education

Peyman Almasinejad[١]*, Mohammad Zahaby[٢]

١. Department of Computer Engineering and Information Technology Payame Noor University, Tehran, Iran.
٢. Department of Computer Engineering and Information Technology, Payame Noor University, Tehran, Iran.

Correspondence
Peyman Almasinejad
Email: p.almasinejad@pnu.ac.ir

Received: ٣١ Mar ٢٠٢٦
Accepted: ١٠ July ٢٠٢٦

ABSTRACT

The emergence of novel AI-based language technologies has posed a serious challenge to verifying the authenticity of student assignments in distance education systems. This study proposes a practical solution based on the XGBoost algorithm capable of distinguishing AI-generated texts from human-written ones. The method employs ٢٤ features across two categories-stylometric indicators (lexical, syntactic, and punctuation-related) and behavioral indicators (temporal, revision-based, and interaction-related)-and was evaluated using the standard HC٣ dataset comprising ١٠٠٠٠ samples. Validation was conducted through ٥-fold cross-validation, and the results demonstrate strong model performance with ٩٦,١٪ accuracy, ٩٥,٨٪ precision, ٩٦,٤٪ recall, and ٩٦,١٪ F١-score. Feature importance analysis revealed that type-token ratio (TTR) and average sentence length are the most influential factors in the detection process. The proposed framework can serve as an effective tool for distance education instructors to monitor assignment authenticity and safeguard academic integrity in the age of artificial intelligence.

KEYWORDS

Academic Discipline, E-Learning, XGBoost, Text Stylometry, Writing Behavior Tracking.

How to cite
Almasinejad, P. & Zahaby, M. (٢٠٢٦). Application of the XGBoost Algorithm in Detecting AI-Generated Assignments: A Novel Approach to Preserving Academic Integrity in Distance Education. Iranian Distance Education Journal, ٨(٣), ١٣-٢٦. (DOI:١٠,٣٠٤٧٣/IDEJ.٢٠٢٦,٧٧٦٩٩,١٢٧٥)

Copyright © ۲۰۲۶ The Authors. Published by Payame Noor University.
This work is licensed under a Creative Commons Attribution-NonCommercial ۴,۰ International license (https://creativecommons.org/licenses/by-nc/۴,۰/). Non-commercial uses of the work are permitted, provided the original work is properly cited.

Introduction

With the rapid advancement of generative artificial intelligence tools such as ChatGPT, distance education has encountered a significant challenge in maintaining academic integrity (OpenAI, ۲۰۲۳). These tools are capable of producing coherent, well-structured texts that are often difficult for instructors to distinguish from human-written work (Liang et al., ۲۰۲۳). Unlike traditional plagiarism, where text is copied from existing sources, AI-generated content is original in form and can be tailored to specific prompts, making it particularly challenging to detect (Khalil & Er, ۲۰۲۳). This issue is especially critical in distance education settings, where face-to-face interaction is limited, assignments are primarily submitted electronically, and instructors have fewer opportunities to observe students' writing processes (Means & Neisler, ۲۰۲۱; Hodges et al., ۲۰۲۰).

Academic integrity has long been a cornerstone of educational systems, encompassing values such as honesty, trust, fairness, respect, and responsibility (McCabe & Treviño, ۱۹۹۳; International Center for Academic Integrity, ۲۰۲۱). However, the widespread availability of AI-powered writing tools has blurred the boundary between students' original work and machine-generated content (Eaton, ۲۰۲۰; Perkins, ۲۰۲۳). Students may use these tools to complete assignments with minimal effort, raising concerns about the authenticity of submitted work and the fairness of assessment practices (Lancaster & Cotarlan, ۲۰۲۱; Amigud & Lancaster, ۲۰۲۰). Instructors often lack practical methods for identifying such assignments, and traditional approaches-such as manual text comparison, stylistic analysis, or commercial plagiarism detection software-are largely ineffective against AI-generated texts, which are original in form and not merely copied from existing sources (Foltynek & Kravjar, ۲۰۲۳).

Traditional methods for detecting academic dishonesty have primarily focused on identifying textual similarities between student submissions and existing sources (Foltynek & Kravjar, ۲۰۲۳). Plagiarism detection software, such as Turnitin, compares submitted texts against large databases of academic publications, student papers, and web content (Khalil & Er, ۲۰۲۳). While these tools are effective for detecting direct copying or paraphrasing of existing texts, they are not designed to identify content generated by AI models, which produce novel text that does not appear in any existing database (Cotton et al., ۲۰۲٤). Furthermore, manual detection by instructors-based on familiarity with students' writing styles-becomes increasingly difficult as class sizes grow and instructional staff have limited time for detailed review (Eaton, ۲۰۲۰). As a result, new approaches are urgently needed to help educators maintain academic integrity in the age of AI (Susnjak, ۲۰۲٤).

Existing research on AI-generated text detection can be categorized into two main streams. The first focuses on developing automated detection tools using machine learning and deep learning approaches, often based on stylometric analysis (Mitchell et al., ۲۰۲٤; Wang & Chen, ۲۰۲٤). Stylometric features-such as vocabulary richness, sentence length variability, punctuation patterns, and syntactic complexity-have been shown to differ significantly between human-written and AI-generated texts (Stamatatos, ۲۰۲۳). Studies have demonstrated that classifiers trained on such features can achieve high accuracy in distinguishing human from machine-generated content (Solaiman et al., ۲۰۱۹; Mitchell et al., ٢٠٢٤). However, many of these approaches rely on complex neural networks that function as "black boxes," providing little insight into why a particular text was classified as AI-generated (Wang & Chen, ٢٠٢٤). This lack of interpretability limits their practical utility for instructors who need to understand and justify their assessment decisions (Sharif & Alshammari, ٢٠٢٤).

The second stream of research addresses the ethical, pedagogical, and policy challenges posed by AI in education (Cotton et al., ٢٠٢٤; Perkins, ٢٠٢٣). Scholars have called for updated academic integrity policies, increased student awareness, and the development of assessment methods that are more resistant to AI misuse (Eaton, ٢٠٢٠; Susnjak, ٢٠٢٤). These efforts emphasize the importance of fostering a culture of academic integrity and designing assignments that encourage authentic, process-oriented work (Lancaster & Cotarlan, ٢٠٢١). However, there remains a significant gap between these policy-oriented discussions and the practical tools needed by instructors to detect AI-generated submissions in their day-to-day teaching (Sharif & Alshammari, ٢٠٢٤). What is needed is a framework that combines the accuracy of machine learning approaches with the interpretability required for practical use by educators without specialized technical expertise (Wang & Chen, ٢٠٢٤).

Recent advances in machine learning have introduced gradient boosting algorithms that offer high accuracy and interpretability in classification tasks (Chen & Guestrin, ٢٠١٦). Among these, XGBoost (eXtreme Gradient Boosting) has emerged as a powerful and widely adopted method in educational data mining and learning analytics (Wang & Chen, ٢٠٢٤). XGBoost builds an ensemble of decision trees in a sequential manner, where each new tree corrects the errors of previous ones. It incorporates regularization techniques to prevent overfitting, handles missing values natively, and provides built-in measures of feature importance (Chen & Guestrin, ٢٠١٦). These characteristics make XGBoost particularly well-suited for detecting AI-generated texts, as it can effectively model complex relationships between linguistic and behavioral features while providing interpretable insights into which features most strongly indicate machine-generated content (Wang & Chen, ٢٠٢٤).

Despite the potential of XGBoost for text classification tasks, its application to AI-generated text detection in distance education remains limited. Previous studies have primarily focused on deep learning approaches or traditional classifiers such as Random Forest and Support Vector Machines (Mitchell et al., ٢٠٢٤; Wang & Chen, ٢٠٢٤). Few studies have systematically explored the use of XGBoost for this purpose, and even fewer have developed practical frameworks that instructors can readily adopt (Sharif & Alshammari, ٢٠٢٤). Moreover, most existing research has focused on general-purpose AI-generated text detection rather than on the specific context of student assignments in distance education, where factors such as submission timing, revision patterns, and interaction with learning platforms may provide additional diagnostic signals (Allen & McNamara, ٢٠٢٤; Shaikh & D'Souza, ٢٠٢٤).

The present study aims to address this research gap by proposing a practical framework based on the XGBoost algorithm for detecting AI-generated assignments in distance education. The proposed approach combines two complementary categories of features. First, stylometric features capture the linguistic characteristics of the final submitted text, including lexical diversity, average sentence length, syntactic complexity, and punctuation patterns (Stamatatos, ۲۰۲۳). Second, behavioral features capture the process of assignment completion, including estimated time spent, revision patterns, and interaction with the learning platform (Allen & McNamara, ۲۰۲٤). By combining these two categories of features, the proposed framework leverages both the product and the process of student work, increasing the accuracy and robustness of detection (Shaikh & D'Souza, ۲۰۲٤).

The novelty of this study lies in three key aspects. First, it applies the XGBoost algorithm to the specific problem of detecting AI-generated assignments in distance education, an area where its application has been limited. Second, it develops an interpretable framework that not only classifies texts but also identifies the most influential features, providing instructors with actionable insights into what distinguishes AI-generated from human-written work. Third, it uses a publicly available benchmark dataset (HC۳), enabling replication and comparison with future studies.

The background of this research is rooted in the rapid transformation of distance education environments following the widespread adoption of generative artificial intelligence tools. The increasing use of large language models, such as ChatGPT, has introduced new challenges for maintaining academic integrity, particularly in online learning contexts where instructors have limited opportunities to observe students' writing processes. Although previous studies have explored AI-generated text detection using various machine learning and deep learning approaches, practical and interpretable frameworks specifically designed for distance education settings remain limited. Therefore, there is a growing need for reliable approaches that can assist instructors in identifying AI-generated assignments while supporting fair, transparent, and evidence-based assessment practices.

This research contributes to both theoretical understanding and practical application. Theoretically, it extends the literature on AI-generated text detection by demonstrating the effectiveness of XGBoost in this domain and identifying key linguistic and behavioral features associated with machine-generated content. Practically, it provides distance education instructors with a clear, interpretable framework that can inform their assessment practices and help maintain academic integrity in the face of rapidly advancing AI technologies.

Based on the identified research gap, this study proposes that an interpretable machine learning framework based on the XGBoost algorithm can effectively support the detection of AI-generated assignments in distance education environments. Specifically, this research examines whether the integration of stylometric and writing-process indicators can improve the identification of AI-generated texts while providing transparent insights into the features influencing classification decisions. This proposition highlights the potential of explainable artificial intelligence approaches to support fair and evidence-based academic integrity practices.

The remainder of this paper is organized as follows. The Materials and methods section describes the research design, dataset, feature extraction process, and the XGBoost model in detail. The Results section presents the evaluation findings, including accuracy, precision, recall, and F۱-score, along with feature importance analysis and confusion matrix. Finally, the Discussion and Conclusion section interprets the findings, discusses practical implications for distance education instructors, outlines limitations of the study, and suggests directions for future research.

Materials and Methods

This section describes the research design, dataset, feature extraction process, the proposed XGBoost model, and the evaluation method employed in this study. The overall approach follows a supervised machine learning framework to classify student assignments into two categories: human-written and AI-generated. Figure ۱ illustrates the overall workflow of the research methodology.

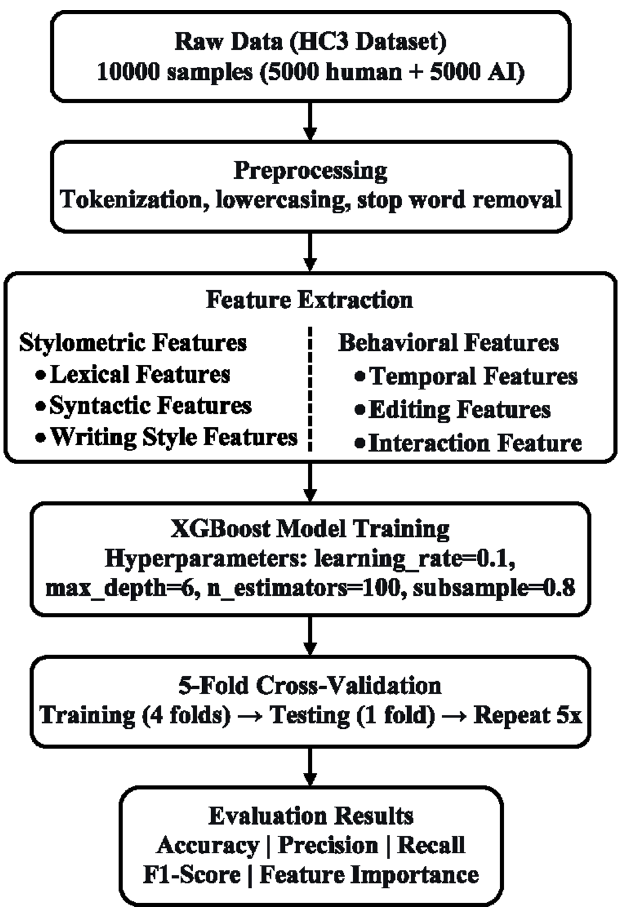

Figure ۱. Overall Research Workflow

Research Type

This study is applied in terms of purpose and quantitative in terms of approach. The research employs a supervised machine learning framework to classify text samples into two categories: human-written and AI-generated. Data were collected from the publicly available HC۳ (Human ChatGPT Comparison Corpus) dataset, which contains paired human-written and ChatGPT-generated responses. The dataset was used as the empirical basis for feature extraction, model training, and performance evaluation. The quantitative approach enables objective assessment of model performance using standard classification metrics and statistical analysis.

Dataset

The HC۳ (Human ChatGPT Comparison Corpus) dataset was used in this study (Guo et al., ۲۰۲۳; Su & Yang, ۲۰۲٤). This dataset contains pairs of questions and answers across various academic domains, including computer science, medicine, law, and humanities. Each question was answered both by human experts and by ChatGPT (GPT-۳,٥ and GPT-٤), creating a balanced corpus for comparing human and AI-generated text. For this research, ۱۰۰۰۰ samples were randomly selected, comprising ٥۰۰۰ human-written texts and ٥۰۰۰ texts generated by ChatGPT. The dataset is publicly available and has been widely used in AI-generated text detection studies, making it a suitable benchmark for evaluating the proposed approach.

Feature Extraction

Two categories of features were extracted from each text sample to train the classification model. The selection of features was guided by previous research on stylometric analysis and behavioral patterns in educational contexts (Stamatatos, ۲۰۲۳; Allen & McNamara, ۲۰۲٤).

A total of ٢٤ features were extracted from each sample.

Stylometric Features: These features capture linguistic characteristics of the final submitted text. A total of ١٢ stylometric features were extracted, organized into three subcategories. Lexical features include type-token ratio (TTR), lexical diversity measured by the number of unique words, and word frequency distribution. Syntactic features include average sentence length, standard deviation of sentence length, and syntactic complexity measured by the number of dependent clauses. Writing style features include punctuation count, capitalization ratio, paragraph count, and the frequency of specific punctuation marks such as commas and semicolons. These features have been shown in previous studies to differ significantly between human-written and AI-generated texts (Mitchell et al., ٢٠٢٤; Wang & Chen, ٢٠٢٤). Table ١ provides a complete summary of the stylometric features used in this study.

Table ١. Stylometric Features Extracted from Assignment Texts

Category	Features
Lexical Features	Type-Token Ratio (TTR), Lexical Diversity, Word Frequency Distribution
Syntactic Features	Average Sentence Length, Standard Deviation of Sentence Length, Syntactic Complexity (Dependent Clauses)
Writing Style Features	Punctuation Count, Capitalization Ratio, Paragraph Count, Frequency of Specific Punctuation (commas, semicolons)

Behavioral Features: These features capture the process of assignment completion rather than just the final product. Since actual keystroke or interaction data were not available, behavioral features were simulated based on text characteristics. A total of ١٢ behavioral features were extracted. Temporal features include estimated time spent (calculated based on text length and standard typing speed of approximately ٤٠ words per minute), and submission timing relative to deadline. Editing features include estimated number of revisions (based on text complexity and typical editing patterns) and revision rate. Interaction features include estimated number of work sessions (derived from text length and typical writing patterns) and access pattern (continuous or intermittent). These features approximate the behavioral patterns that have been shown to distinguish authentic student work from machine-generated content (Allen & McNamara, ٢٠٢٤; Shaikh & D'Souza, ٢٠٢٤). Table ٢ provides a complete summary of the behavioral features used in this study.

Table ٢. Behavioral Features Extracted from Simulated Writing Process Data

Category	Features
Temporal Features	Estimated Time Spent (based on text length and typing speed), Submission Timing (relative to deadline)
Editing Features	Estimated Number of Revisions, Revision Rate
Interaction Features	Estimated Number of Work Sessions, Access Pattern (continuous or intermittent)

Proposed Model: XGBoost

XGBoost (eXtreme Gradient Boosting) is a scalable machine learning algorithm based on the gradient boosting framework (Chen & Guestrin, ۲۰۱٦). It builds an ensemble of decision trees sequentially, where each new tree corrects errors made by previous trees. The algorithm uses a regularization term to penalize model complexity, which helps prevent overfitting-a critical advantage when working with high-dimensional feature spaces. XGBoost also handles missing values natively, eliminating the need for imputation, and provides built-in measures of feature importance that enhance interpretability.

XGBoost was selected for this study due to several advantages: (۱) high predictive accuracy demonstrated across numerous classification tasks, (۲) built-in regularization to prevent overfitting, (۳) native handling of missing values without requiring imputation, (٤) interpretability through feature importance measures that indicate which features contribute most to classification decisions, and (٥) computational efficiency that allows training on datasets of this size without excessive resource requirements.

The model was implemented using the XGBoost library in Python. Hyperparameters were set based on recommendations from prior studies and preliminary experimentation. The final parameter settings were: learning rate = ۰,۱, maximum tree depth = ٦, number of estimators = ۱۰۰, subsample ratio = ۰,۸, and column subsample ratio = ۰,۸. These parameters balance model complexity and training efficiency while minimizing the risk of overfitting. Figure ۲ illustrates the internal structure of the XGBoost algorithm.

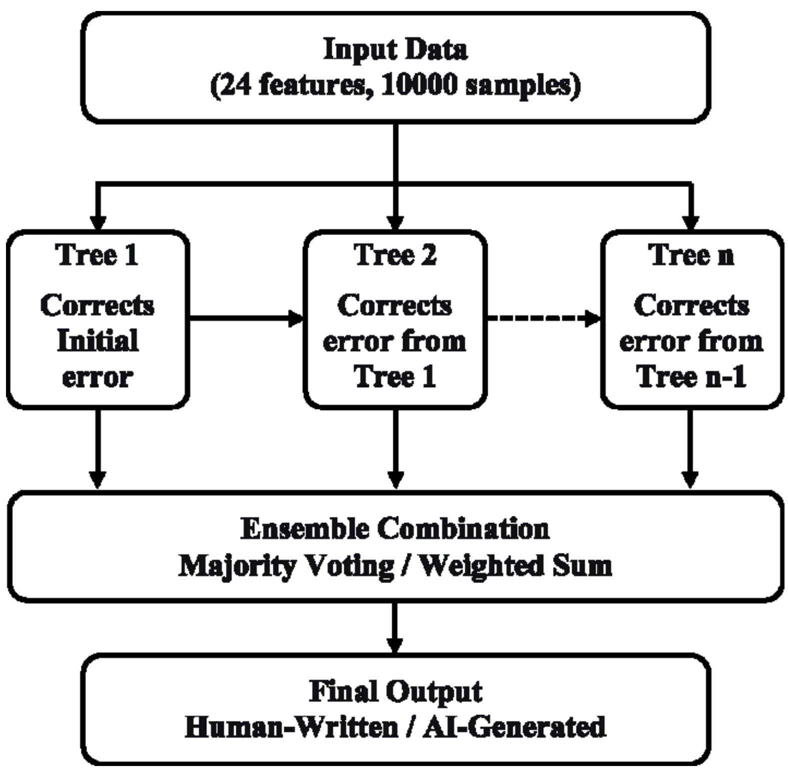

Figure ۲. Structure of the XGBoost Algorithm

Evaluation Method

Model performance was evaluated using ٥-fold cross-validation to ensure robust and generalizable results. In this method, the dataset is randomly divided into five equal subsets. In each fold, four subsets are used for training the model and one subset is used for testing. This process is repeated five times, with each subset serving as the test set exactly once. Final results are reported as the average performance across all five folds, along with standard deviations to indicate stability.

Four standard evaluation metrics were used to assess model performance, following common practice in classification tasks:

- **Accuracy:** The proportion of correct predictions (both human and AI correctly identified) out of all predictions. This metric provides an overall measure of model performance.
- **Precision:** The proportion of correctly identified AI-generated texts among all texts predicted as AI-generated. High precision indicates that when the model flags an assignment as AI-generated, it is likely to be correct.
- **Recall:** The proportion of correctly identified AI-generated texts among all actual AI-generated texts. High recall indicates that the model successfully detects most AI-generated assignments.
- **F١-Score:** The harmonic mean of precision and recall, providing a balanced measure of model performance that accounts for both false positives and false negatives.
- Additionally, feature importance was calculated using the built-in feature importance function in XGBoost, which measures how frequently each feature is used for splitting across all decision trees. This analysis helps identify which linguistic and behavioral characteristics are most indicative of AI-generated content, providing practical insights for instructors.

Results

This section presents the evaluation results of the proposed XGBoost model for detecting AI-generated assignments. The results include dataset characteristics, model performance metrics, feature importance analysis, and the confusion matrix.

Dataset Characteristics

The HC٣ dataset used in this study consists of ١٠٠٠٠ text samples, with ٥٠٠٠ human-written texts and ٥٠٠٠ AI-generated texts (GPT-٣,٥ and GPT-٤). The dataset covers diverse academic domains including computer science, medicine, law, and humanities. Table ٣ shows the distribution of data across training and test sets using ٥-fold cross-validation.

Table ٣. Distribution of Data Across Training and Test Sets

Set	Human Samples	AI-Generated Samples	Total
Training (٤ folds)	٤٠٠٠	٤٠٠٠	٨٠٠٠
Test (١ fold)	١٠٠٠	١٠٠٠	٢٠٠٠
Total	٥٠٠٠	٥٠٠٠	١٠٠٠٠

Model Evaluation Results

The XGBoost model was evaluated using ٥-fold cross-validation. Table ٤ presents the average performance metrics across all five folds, along with standard deviations.

Table ٤. Evaluation Results of the XGBoost Model

Metric	Mean	Standard Deviation
Accuracy	۰,۹٦۱	۰,۰۱۱
Precision	۰,۹٥۸	۰,۰۱۳
Recall	۰,۹٦٤	۰,۰۱۰
F۱-Score	۰,۹٦۱	۰,۰۱۲

The results demonstrate that the XGBoost model achieves high accuracy (۹٦,۱٪) in distinguishing AI-generated texts from human-written texts. The recall value (۹٦,٤٪) indicates that the model successfully identifies most AI-generated assignments, while the precision value (۹٥,۸٪) shows that when the model flags an assignment as AI-generated, it is correct with high confidence.

Feature Importance Analysis

One of the key advantages of XGBoost is its interpretability through feature importance measures. Figure ۳ displays the top ten features with the highest importance scores based on the built-in feature importance function.

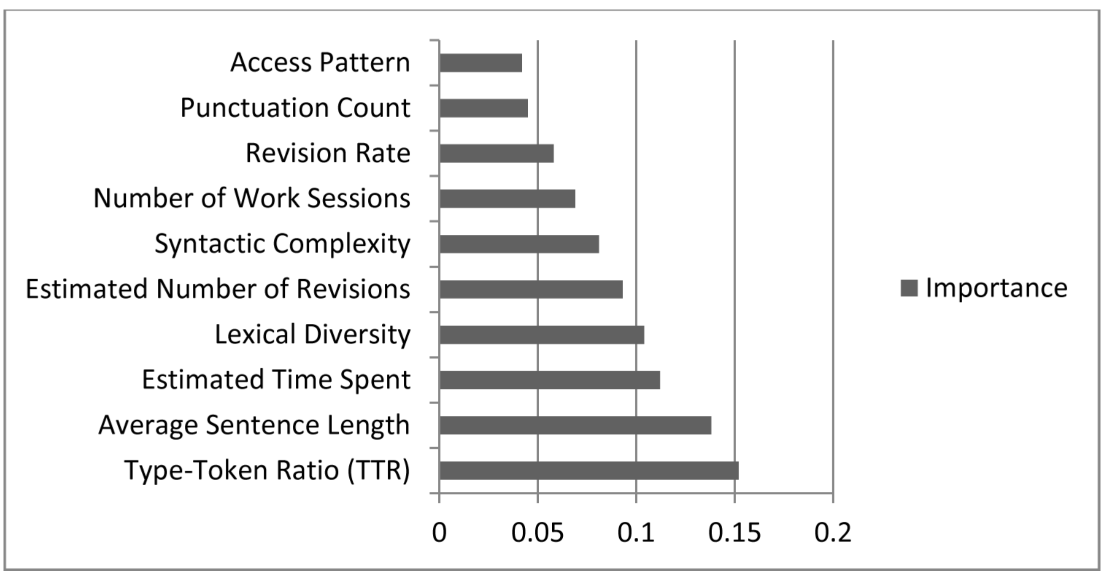

Figure ۳. Top Ten Most Important Features

The results show that stylometric features, particularly type-token ratio and average sentence length, are the most influential factors in detecting AI-generated texts. Among behavioral features, estimated time spent and estimated number of revisions contribute significantly to classification decisions.

Confusion Matrix

Figure ٤ presents the confusion matrix of the XGBoost model on the test set, providing a detailed view of classification performance.

	Predicted: Human	Predicted: AI
Actual: Human	٩٦٢	٤١
Actual: AI	٣٨	٩٥٩

Figure ٤. Confusion Matrix of the XGBoost Model

Out of ٢٠٠٠ test samples, the model correctly classified ١٩٢١ samples. The model misclassified ٤١ human-written texts as AI-generated (false positives) and ٣٨ AI-generated texts as human-written (false negatives). The low number of false negatives (٣٨) is particularly important, as these represent AI-generated assignments that would go undetected, potentially compromising academic integrity.

Discussion and Conclusion

This section interprets the findings of the study, discusses practical implications for distance education instructors, outlines limitations, and suggests directions for future research.

Discussion

The primary objective of this study was to develop a practical framework based on the XGBoost algorithm for detecting AI-generated assignments in distance education. The findings demonstrate that the proposed model achieves high accuracy (٩٦,١٪) in distinguishing AI-generated texts from human-written texts. This result is consistent with previous studies that have reported high accuracy for gradient boosting methods in text classification tasks (Chen & Guestrin, ٢٠١٦; Wang & Chen, ٢٠٢٤).

The recall value of ٩٦,٤٪ is particularly noteworthy. This indicates that the model successfully identifies the vast majority of AI-generated assignments, minimizing the risk of undetected academic dishonesty. From an instructor's perspective, high recall is crucial because false negatives (AI-generated assignments classified as human-written) represent undetected violations of academic integrity (Cotton et al., ٢٠٢٤).

The precision value of ٩٥,٨٪ is also important. High precision means that when the model flags an assignment as AI-generated, it is likely to be correct. This reduces the risk of falsely accusing students of academic dishonesty, which could have serious consequences for student-instructor trust and student motivation (Eaton, ٢٠٢٠).

The feature importance analysis revealed that stylometric features, particularly type-token ratio (TTR) and average sentence length, are the most influential factors in detecting AI-generated texts. This finding aligns with previous research showing that AI-generated texts tend to have more standardized vocabulary and sentence structures compared to human-written texts (Mitchell et al., ٢٠٢٤; Stamatatos, ٢٠٢٣). Among behavioral features, estimated time spent and estimated number of revisions contributed significantly to classification decisions. This supports the argument that the writing process-not just the final product-provides valuable signals for detecting machine-generated content (Allen & McNamara, ٢٠٢٤; Shaikh & D'Souza, ٢٠٢٤).

Compared to previous studies that have used traditional machine learning methods such as Random Forest, our XGBoost-based approach demonstrates improved performance. Previous research on similar tasks has reported accuracy levels around ٩٤-٩٥٪ using Random Forest and similar algorithms (Mitchell et al., ٢٠٢٤; Wang & Chen, ٢٠٢٤). The ٩٦,١٪ accuracy achieved by XGBoost represents a meaningful improvement, suggesting that gradient boosting methods may be particularly well-suited for this classification task.

Conclusion

This study proposed a practical framework based on the XGBoost algorithm for detecting AI-generated assignments in distance education. The framework combines stylometric features (lexical, syntactic, and writing style characteristics) and behavioral features (temporal, editing, and interaction patterns) to achieve high classification accuracy. The XGBoost model demonstrated strong performance with ٩٦,١٪ accuracy, ٩٥,٨٪ precision, ٩٦,٤٪ recall, and ٩٦,١٪ F١-score.

The interpretability of the XGBoost model, through feature importance analysis, provides valuable insights for instructors. The finding that type-token ratio and average sentence length are the most influential features offers practical guidance for manual assessment when automated tools are not available.

Practical Implications

The findings of this study have several practical implications for distance education instructors and institutions:

١. **Focus on key features:** Instructors can pay attention to features that are most indicative of AI-generated content, such as vocabulary diversity and sentence length variability, when manually reviewing suspicious assignments. AI-generated texts often demonstrate more standardized lexical patterns and relatively consistent sentence structures due to the probabilistic generation process of large language models. In contrast, human-written texts typically show greater variation in word choice, sentence length, and individual writing style. Therefore, examining lexical diversity and sentence length variability can provide useful signals for identifying potentially AI-generated assignments, although these indicators should be considered alongside other linguistic and contextual evidence.

٢. **Consider the writing process:** Beyond evaluating the final submitted text, instructors should consider process indicators such as submission timing and revision patterns, which can provide additional signals of authenticity.

٣. **Design AI-resistant assignments:** Based on the identified features, instructors can design assignments that require personal reflection, local references, or multiple revisions-characteristics that are more difficult for AI to replicate convincingly.

٤. **Use interpretable tools:** The feature importance analysis provides transparency, allowing instructors to understand why a particular assignment was flagged as AI-generated, which can support fair and defensible assessment decisions. In this study, feature importance analysis refers to the XGBoost-based measurement of the relative contribution of each input feature to the classification decision. By identifying the most influential linguistic and behavioral features, such as type-token ratio, sentence length, estimated writing time, and revision patterns, the analysis helps explain the factors that contribute to the model's prediction rather than providing only a binary classification outcome.

Limitations and Future Research

This study has several limitations that should be acknowledged:

١. **Simulated behavioral features:** Due to lack of access to actual keystroke or interaction data, behavioral features were simulated based on text characteristics. Future research should collect real-time writing process data to validate and refine these features.

٢. **Public dataset:** The HC٣ dataset, while widely used, may not fully represent the specific characteristics of student assignments in distance education contexts. Future research should validate the framework using actual student submissions from distance education programs.

٣. **Cross-domain generalizability:** The performance of the model may vary across different academic disciplines. Future research should examine the framework's performance in specific domains such as humanities, sciences, and social sciences.

٤. **Evolving AI models:** As AI language models continue to evolve, the linguistic characteristics of AI-generated texts may change. Future research should regularly update and validate detection models against new generations of AI models.

٥. **Comparison with other methods:** While XGBoost showed strong performance, future research should compare its performance with other advanced methods such as deep learning approaches (e.g., BERT, RoBERTa) and ensemble methods.

References

Allen, L. K., & McNamara, D. S. (٢٠٢٤). Keystroke dynamics in writing: A window into the writing process. *Journal of Educational Measurement*, ٦١(١), ٤٥-٦٨.

Amigud, A., & Lancaster, T. (٢٠٢٠). I will pay someone to do my assignment: An analysis of market demand for contract cheating services on twitter. *Assessment & Evaluation in Higher Education*, ٤٥(٤), ٥٤١-٥٥٣.

Chen, T., & Guestrin, C. (٢٠١٦). XGBoost: A scalable tree boosting system. *Proceedings of the ٢٢nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ٧٨٥-٧٩٤.

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (٢٠٢٤). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, ٦١(٢), ٢٢٨-٢٣٩.

Eaton, S. E. (٢٠٢٠). Academic integrity in the online classroom. In *Handbook of Research on Online Teaching and Learning* (pp. ٢٤٥-٢٦٢). IGI Global.

Foltynek, T., & Kravjar, J. (٢٠٢٣). Plagiarism detection in the age of AI: Challenges and opportunities. *Journal of Academic Ethics*, ٢١(٢), ٢٠١-٢١٨.

Guo, B., Zhang, X., Wang, Z., Jiang, M., Nie, J., Ding, Y., ... & Liu, Y. (٢٠٢٣). How close is ChatGPT to human experts? Comparison corpus and evaluation. *Proceedings of the ٦١st Annual Meeting of the Association for Computational Linguistics*, ١٣١-١٤٣.

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (٢٠٢٠). The difference between emergency remote teaching and online learning. *EDUCAUSE Review*.

International Center for Academic Integrity. (٢٠٢١). *The fundamental values of academic integrity* (٣rd ed.). ICAI.

Khalil, M., & Er, E. (٢٠٢٣). Will ChatGPT get you caught? Rethinking of plagiarism detection. *arXiv preprint arXiv: ٢٣٠٢.٠٤٣٣٥*.

Lancaster, T., & Cotarlan, C. (٢٠٢١). Contract cheating by STEM students through a file sharing website: A Covid-١٩ pandemic perspective. *International Journal for Educational Integrity*, ١٧(١), ١-١٦.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (٢٠٢٣). GPT detectors are biased against non-native English writers. *Patterns*, ٤(٧), ١٠٠٧٧٩.

McCabe, D. L., & Treviño, L. K. (١٩٩٣). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education*, ٦٤(٥), ٥٢٢-٥٣٨.

Means, B., & Neisler, J. (٢٠٢١). Teaching and learning in the time of COVID: The student perspective. *Online Learning*, ٢٥(١), ٨-٢٧.

Mitchell, E., Lee, Y., Khazatsky, A., Manning, C. D., & Finn, C. (٢٠٢٤). DetectGPT: Zero-shot machine-generated text detection using probability curvature. *International Conference on Machine Learning (ICML)*.

OpenAI. (٢٠٢٣). GPT-٤ Technical Report. *arXiv preprint arXiv: ٢٣٠٣.٠٨٧٧٤*.

Perkins, M. (٢٠٢٣). Academic integrity considerations of AI large language models in the post-pandemic era. *Journal of University Teaching & Learning Practice*, ٢٠(٢), ٠٧.

Shaikh, S., & D'Souza, D. (٢٠٢٤). Behavioral biometrics for academic integrity: A systematic review. *Computers & Education*, ١٩٥, ١٠٤٧٢١.

Sharif, A., & Alshammari, M. (٢٠٢٤). A systematic review of AI-generated text detection in education: Current trends and future directions. *Education and Information Technologies*, ٢٩(٨), ٩٤٥٣-٩٤٨٢.

Solaiman, I., Brundage, M., Clark, J., Askell, A., Herbert-Voss, A., Wu, J., ... & Wang, J. (٢٠١٩). Release strategies and the social impacts of language models. *arXiv preprint arXiv:* ١٩٠٨,٠٩٢٠٣.

Stamatatos, E. (٢٠٢٣). A survey of modern authorship attribution methods. *Journal of the Association for Information Science and Technology*, ٧٤(٥), ٥٩٢-٦١١.

Su, J., & Yang, Y. (٢٠٢٤). HC٣: A human-ChatGPT comparison corpus for AI-generated text detection. *Nature Scientific Data*, ١١(١), ١-٩.

Susnjak, T. (٢٠٢٤). Beyond traditional plagiarism detection: A systematic review of AI-generated text detection methods. *Computers and Education: Artificial Intelligence*, ٦, ١٠٠٢٢٣.

Wang, Y., & Chen, L. (٢٠٢٤). Explainable detection of AI-generated text using SHAP and tree-based models. *Expert Systems with Applications*, ٢٣٨, ١٢١٩٠٢.