PLAGIARISM DETECTION IN SUBMITTED ONLINE ASSIGNMENTS - A COMPARATIVE PERFORMANCE ANALYSIS OF THREE MACHINE LEARNING ALGORITHMS

ODULAJA Godwin Oluseyi.
Department of Computer and Information Sciences,
College of Science and Information Technology,
Federal University of Education Ijagun, Ogun State, Nigeria
goddyseyi@gmail.com | (+234) 8052064987

OGUNSANWO Gbenga .O.
Department of Computer and Information Sciences,
College of Science and Information Technology,
Federal University of Education Ijagun, Ogun State, Nigeria
ogunsanwogo@tasued.edu.ng

OKUSANYA Adedoyin O.
Department of Entrepreneurial Studies,
Tai Solarin Federal University of Education
Ijagun, Ogun State, Nigeria
okusanyao@tasued.edu.ng

&

OGUNBANJO Remilekun O.
Department of Computer and Information Sciences,
College of Science and Information Technology,
Federal University of Education Ijagun, Ogun State, Nigeria
ogunbanjoro@tasued.edu.ng

Abstract

Literature showed that plagiarism has globally become a malignant and destructive practice in tertiary institutions. Artificial Intelligence (AI) tools has facilitated high-tech plagiarism, and consequently undermined authenticity of academic works; bringing credibility of academic programs, value of degrees, and the skills of graduates under threat. This study developed an AI model for performance evaluation of machine learning algorithms in detecting plagiarism in online assignment submission. A dataset of 200 submissions obtained from A Corpus of Plagiarized Short Answers (ACOPSA)) and local student assignment submissions was trained with Random Forest, XGBoost, and K-Nearest Neighbors machine learning algorithms in the model. Precision, Recall, F1 score, Accuracy, and ROC-AUC were employed for comparative performance assessment of the algoritms,. Findings showed that Random Forest (Accuracy: 0.857, Precision: 0.858, and F1 Score: 0.854) achieved the best overall performance, followed by XGBoost (Accuracy: 0.800, Precision: 0.820, and F1 Score: 0.795) while KNN yielded the least performance (Accuracy: 0.579, Precision: 0.591, and F1 Score: 0.546). Random Forest demonstrated superior capability in detecting paraphrased and heavily disguised plagiarism. For academic integrity sustenance in all academic institutions, the researchers recommended development of hybrid models that concurrently detects lexical and semantic matches, and enforcement of ethical use of AI.

Keywords: Plagiarism Detection, Online Assignment, Academic Integrity, Machine Learning Algorithms, Hybrid Models.

Introduction

Increasing availability and accessibility to digital tools and platforms occasioned by the Internet, has made plagiarism a pervasive issue in academic environments, widespread, prevalent and difficult to detect. The fact that students admitting to cheating and copying academic works are on the increase is worrisome and this worrying trend is largely driven by the ease of copying and pasting from online sources (Oravec, 2023). In addition to traditional forms of plagiarism, such as directly copying text, students now have access to powerful AI-based tools that can paraphrase content and manipulate it to avoid detection by conventional plagiarism checkers (Kumar, 2021). With these advances, students can now produce academic work that appears original but is, in fact, generated by AI systems or heavily paraphrased from external sources.

Traditional plagiarism detection tools like Turnitin developed to catch instances of direct copying often struggle with more sophisticated forms of plagiarism, such as paraphrasing or using AI tools to alter the content to evade detection. This has created a significant gap in the effectiveness of these tools, as students increasingly use paraphrasing tools powered by AI to avoid detection (Kumar, 2021). As a result, educational institutions are facing significant challenges in upholding academic integrity.

Furthermore, the rapid development of generative AI technologies, such as OpenAI's ChatGPT, Deepseek, Google's Bard, and Amazon's Bedrock, has introduced new complexities to the issue of academic cheating. These tools are capable of producing high-quality content that can easily be submitted as assignments or exam answers without students contributing any original thought or effort. These tools have become so sophisticated that their outputs can be difficult to distinguish from student-generated work, further complicating the task of detecting academic dishonesty (Huang, 2023). This "AI plagiarism," poses a serious challenge to the integrity of academic assessments (Gillard & Rorabaugh, 2023). The misuse of these AI tools for academic cheating undermines the value of educational qualifications and can have long-term consequences for both institutions and students alike.

The limitations of existing plagiarism detection systems have become increasingly apparent in the face of these emerging technologies. While tools like Turnitin and Grammarly are widely used, they primarily focus on comparing text at a surface level, without truly understanding the underlying meaning of the content(Oravec, 2023). The inability to detect AI-generated text, paraphrased content, or subtle changes in language has made these systems inadequate for addressing modern plagiarism challenges (Dahmen Kayaalp, Ollivier, Pareek, Hirschmann, Karlsson, & Winkler, 2023). As AI technology continues to evolve, there is also a potential for more sophisticated forms of academic dishonesty, making it necessary for educational institutions to adopt more advanced solutions.

Literature Review

Several plagiarism checkers have been developed over the years to mitigate plagiarism, such as PlagAware, 2021; Check for plagiarism, 2021; KIT, 2021; Turnitin, 2021; Blackboard, 2021. Each, with its peculiar limitations. Moreover, plagiarizers are becoming negatively "smarter" and can outsmart these systems. Rearranging and paraphrasing content could successfully trick some of these plagiarism checkers. In addition, there are numerous free online paraphrasing tools powered by Artificial Intelligence (AI), that are able to evade many plagiarism detectors. These tools modify "stolen contents" to such a degree as to evade even the most advanced copy content scanning software (Kumar, 2021). As AI tools continue to evolve and provide more sophisticated capabilities for content generation, there is a pressing need for an advanced, AI-based system to effectively detect and prevent plagiarism and cheating in online assignments. Hence, this study sought to develop a machine-learning AI model for detecting plagiarism and cheating in online assignments and use it to compare the performance of three AI-based machine learning algorithms in detecting plagiarism and cheating in online assignments.

Plagiarism Detection

Plagiarism detection tools have advanced significantly over the years, particularly in detecting verbatim or near-verbatim copying (AmberBlog, 2025). These tools, largely based on word-level fingerprinting or string-matching algorithms, are highly effective for identifying direct plagiarism, but they often struggle with semantic plagiarism, where similar ideas are conveyed using different words or paraphrasing techniques.

Word-Based Versus Semantic Based Detections

Traditional detection systems primarily rely on surface-level lexical similarity. They identify overlaps in word sequences, phrases, or sentence structures (Singh & Kumar, 2022). However, when content is paraphrased—especially using different syntactic forms or vocabulary—the similarity score diminishes even though the semantic meaning remains intact. For example, the original sentence: *"The house belongs to him" could be p*araphrased as: *"That man owns the bungalow" or as "The building is Mr. Raphael's."*

While these sentences convey the same meaning, they have no common lexical tokens. It is therefore difficult for conventional plagiarism detection systems to detect their similarities.

However, integration of deep learning models, natural language processing (NLP) and semantic networks has promoted semantic detection, a new shift in plagiarism detection paradigm from traditional approach that focuses on word-level lexical comparisons, as found in Grammarly and Turnitin, to meaning-level comparisons that focuses on what the intent behind the text is (Chen & Raji, 2023).

However, semantic comparison however is costly computationally. Currently, it is not feasible to carry out deep semantic analysis of several billions of literary works, documents or online resources by brute-force approach. Consequently, assignment of vector or numerical values to documents had been proposed by researchers in order to allow for a more effective large-scale comparison for similarity. This does not require pair-wise comparisons directly. (Ahmad et al., 2024).

Recent works such as SemAntiDetect and ConceptMatch have used ontological graphs and contextual embedding for selected academic fields to advance Heinrich and Maurer's (2000) proposed framework that suggested employing domain-specific ontologies, to detect equivalence between conceptually similar texts (Zhao & Mensah, 2025).

Vector Space Models (VSMs) for Similarity Detection

Vector Space Model (VSM) that uses term frequencies and semantic embedding to convert texts into dimensional vector equivalence is gaining ground in recent times among researchers (Nakpih, 2024; Turney & Pantel, 2025). VSM uses cosine similarity to calculate document similarity by measuring angles between vectors in the space for the documents. The bigger the angle, the less similar are the documents and vice versa. With its focus on semantic rather than lexical similarities, context-conscious VSM plagiarism detection tools like BERT, WORD2Vec and SBERT proved to be more promising in identifying recycled idea, paraphrased content or cross-lingual academic dishonesty (Tariq et al., 2022).

Although computational costs and the need to be domain specific remain a challenge, it is apparent that focus in literature is shifting from traditional less effective lexis (word) based to semantics based plagiarism detection tools for academic integrity preservation in literary works.

Performance Limitations of Existing Plagiarism Detection Tools

The major challenges limiting performance of existing plagiarism detection tools are:

Use of synonyms and paraphrasing - many detection tools could not effectively detect these in literary works. Besides, there are several AI tools that can use word synonyms to restructure and automate document rewriting for splagiarists while retaining the original meaning. Traditional systems that rely on lexical or string-based comparison consequently fail to detect such skillful AI powered plagiarism. Especially worrisome is when the paraphrasing of a document is done on a large scale, even models like GPT and BERT that rely on semantic similarity detection could not measure up as they still require scalability and precision optimization (Zihang, Boqing & Liqiang, 2025).

Offline Contents - when documents whose sources are not available online (such as unpublished dissertations, old textbooks, or physical journals) are plagiarized, plagiarism detection tools like Copyscape and Turnitin who, rely soley on access to digitized and indexed databases become ineffective. Ditto to documents in restricted access domains (Zihang, Boqing & Liqiang, 2025).

Cross-Language Plagiarism – this is now a major concern in the academia. When intellectual contents are translated to another language only to be attributed to another author different from the original author is a worrisome case of cross-language plagiarism (Yilmaz & Fernбndez, 2024). Monolingual detection systems have often failed woefully in such cases as they failed to detect contents hat are semantically the same but presented in another languages.

Addressing the Challenges of Existing Plagiarism Detection Tools

Efforts of notable organisations like Google and Internet Archive in digitizing and indexing vast data bases and knowledge repositories had proved helpful in extending the reach of available intellectual works' similarity detection tools, and thus reducing this challenge to some extent. Intent and semantic based detection tools like SBERT and search engines like ConceptNet that employ the use of NLP (Natural Language Processing) and compare intent and concept in paraphrased documents rather than word similarity are equally promising in overcoming plagiarism (Chen & Raji, 2023). These systems, are however limited by being computationally intensive with respect to needed resources and are therefore not yet scalable enough for real time deployment and analysis, thus limiting applicability of these innovative efforts. Consequently, intuitive manual skill of human evaluators, editors, and reviewers are still indispensable in defending and upholding academic integrity. Besides, some levels of plagiarism such as idea theft definitely require manual human contextual comprehension and interpretation regardless of the supports from automated systems.

Nwohiri, Joda, and Ajayi (2023) confirmed the increasing challenge of plagiarism in academia and highlighted the limitations of existing plagiarism detection tools in their study titled "AI-powered plagiarism detection: leveraging forensic linguistics and natural language processing". Their study proposed development of AI-driven plagiarism detection system that could crawl the web to index articles, generate levels of similarities between documents among other things by applying NLP and forensic linguistics to assess plagiarism levels. They further suggested integration of AI and forensic linguistics into plagiarism detection to enhance detection efficiency.

Oravec (2023), in the article: "Artificial Intelligence - implications for academic cheating: expanding the dimensions of responsible human-ai collaboration with ChatGPT and Bard," explores the ethical concerns surrounding AI-generated content in higher education. The study examines the technological arms race between cheating detection systems and students utilizing AI tools for academic dishonesty. The research identifies how AI-powered methods, including facial recognition and watermarking, can be used to curb cheating while also promoting responsible AI usage. The study concludes that instead of solely focusing on punitive measures, educators should guide students in ethically integrating AI into their academic work, preparing them for a collaborative, AI-driven future.

The systematic review by Sozon, Alkharabsheh, Fong, and Chuan (2024) titled: "Cheating and plagiarism in higher education institutions: a literature review" systematically examines factors contributing to academic dishonesty. The study identifies key drivers of plagiarism, such as academic pressure, lack of integrity awareness, outdated honor codes, and the misuse of AI tools. The findings indicate that cheating and plagiarism are influenced by individual, social, cultural, institutional, and technological factors. The study recommends the establishment of ethical and moral development tutorials, revision of honor codes to incorporate AI tools, and the development of plagiarism detection software to improve students' academic writing and paraphrasing skills. The review concludes that a multi-faceted approach, involving policy changes and

technological advancements, is essential for combating academic dishonesty effectively.

Ibrahim (2023), in the article: "Using AI-based detectors to control AI-assisted plagiarism in essay writing: the terminator versus the machines," investigates the effectiveness of AI-driven classifiers in detecting AI-generated texts in ESL compositions. The study examines the reliability of RoBERT a-based classifiers in identifying machine-generated texts, analyzing a dataset of 240 human-written and AI-generated essays. The findings reveal that while these classifiers can detect AI-generated content, their accuracy is inconsistent. The study concludes that while AI-based detectors provide promising solutions for academic integrity in ESL writing, further refinements are needed to enhance their reliability in detecting AI-assisted plagiarism.

Vani and Gupta (2016) argue that intelligent techniques for detecting high obfuscations are still in their infancy, and most of the available online, standalone, and web-based tools are unable to detect complex manipulations.

Foltynek et al. (2020) evaluated the performance of 15 plagiarism checkers from both the coverage and usability perspectives. The study analyzed texts in eight different languages, including Czech, English, German, Italian, Latvian, Slovak, Spanish, and Turkish, with Wikipedia, online publications, and academic theses serving as the primary sources. It was concluded that better results were obtained for major languages compared to minor ones, that the source of the document significantly influenced the performance of these checkers, and that plagiarism from single sources was more difficult to detect than that from multiple sources.

Objectives

This study compared performance of three machine learning algorithms in detecting plagiarism and cheating in online assignments.

Methodology

This study adopted a design and development research methodology model. It involves system modeling, development, testing, and evaluation of its performance in detecting textual plagiarism.

Figure 1 shows the architectural design of the model..

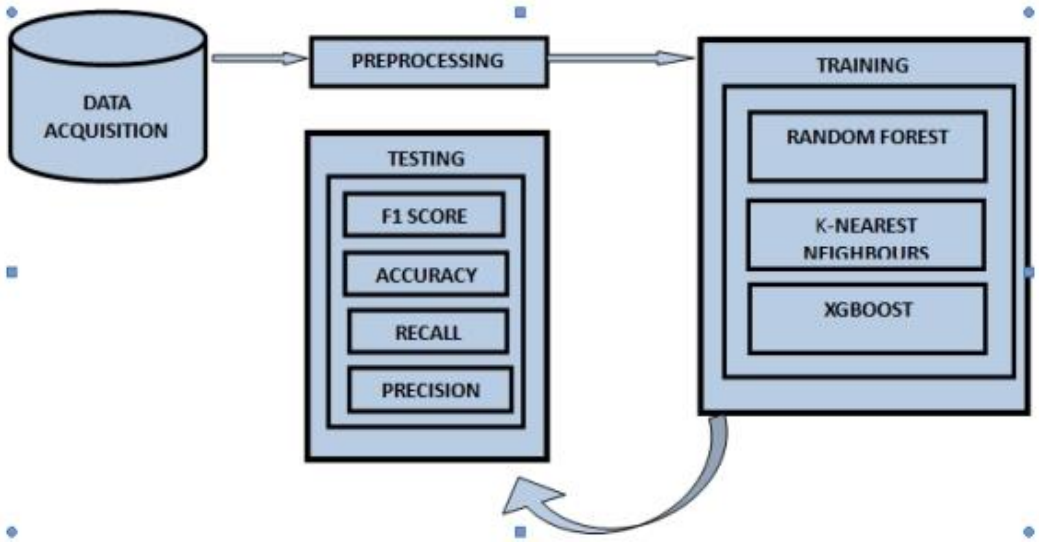

Figure 1 Plagiarism Detection Model's Architectural Design

Implementation

Data Collection

The primary dataset used for training and evaluating the AI-based plagiarism detection model was obtained from the A Corpus of Plagiarized Short Answers (ACOPSA). This dataset is publicly available and has been specifically designed for research in automatic plagiarism detection (see table 1).

Table 1: Snippet of input Dataset for ACOPSA

	File	Group	Person	Task	Category	Native Eng	Knowledge	Difficulty
2	g0pA_task	0	A	a	non	native	1	1
3	g0pA_task	0	A	b	cut	native	4	3
4	g0pA_task	0	A	c	light	native	5	3
5	g0pA_task	0	A	d	heavy	native	3	4
6	g0pA_task	0	A	e	non	native	4	3
7	g0pB_task	0	B	a	non	native	2	1
8	g0pB_task	0	B	b	non	native	3	3
9	g0pB_task	0	B	c	cut	native	5	3
10	g0pB_task	0	B	d	light	native	2	2
11	g0pB_task	0	B	e	heavy	native	4	3
12	g0pC_task	0	C	a	heavy	native	4	3
13	g0pC_task	0	C	b	non	native	3	3
14	g0pC_task	0	C	c	non	native	2	4
15	g0pC_task	0	C	d	cut	native	1	5
16	g0pC_task	0	C	e	light	native	2	2
17	g0pD_task	0	D	a	cut	non-native	1	1
18	g0pD_task	0	D	b	light	non-native	2	2
19	g0pD_task	0	D	c	heavy	non-native	3	3
20	g0pD_task	0	D	d	non	non-native	2	3
21	g0pD_task	0	D	e	non	non-native	3	3
22	g0pE_task	0	E	a	light	non-native	1	1
23	g0pE_task	0	E	b	heavy	non-native	2	2
24	g0pF_task	0	F	c	non	non-native	2	2

Source: www. Zenodo.org/

The ACOPSA dataset was developed by Barro, B., Tounkara, T., & Franɜois, T., and is accessible via academic repositories such as Zenodo or other relevant open research data platforms. It has been used in several NLP and educational integrity studies due to its carefully annotated plagiarism classes.

Data Quantity and Distribution:

The secondary dataset consists of 100 short answer documents,, evenly distributed across four distinct categories:

1. **Non-Plagiarized** – 25 documents containing original answers written independently by participants.
2. **Light Plagiarism** – 25 documents with minor word substitutions or slight paraphrasing.
3. **Heavy Plagiarism** – 25 documents that are heavily reworded but maintain the original meaning.
4. **Cut-Paste Plagiarism** – 25 documents that contain exact copies from source materials with no modification.

This balanced distribution enables the model to learn and distinguish among varying degrees of textual plagiarism.

Local Content Contribution:

In addition to the ACOPSA dataset, primary data was also sourced locally to improve contextual relevance. This includes 100 sample of real student assignments (e.g., essays, code submissions). These samples were collected from academic institutions under informed consent and anonymity to supplement the model's training dataset with local content, and for model evaluation.

Data Cleaning:

To ensure consistency, prior to training, the dataset was cleansed by removing special characters, unnecessary whitespace, and any HTML tags, standardizing punctuation and casing and by eliminating any duplicate records.

Preprocessing:

At this stage four NLP techniques were applied. Using methods such as TF-IDF or word embeddings, vectorisation was used to convert text into equivalent numerical representations; lemmatisation for word reduction to their basic dictionary form (e.g., "converting" → "convert"); tokenization for splitting each text into individual words or tokens and Stop-word Removal for removing determinants, auxiliary verbs (e.g., "are", "the") and other words that has no contribution to semantic meaning. After preprocessing, the dataset was divided into two sets for training and testing on ratio 80:20 respectively.

Data Analysis and Evaluation Metrics

Performance of machine learning algorithms (Random Forest, XGBoost, and K-Nearest Neighbours (KNN)) evaluated was measured using accuracy, precision, F1 score, confusion matrices, and Receiver Operating Characteristic (ROC) curves with Area Under the Curve (AUC) values. Since plagiarism detection does not have a definitive "correct" answer, evaluation was conducted using both qualitative and quantitative measures. These measures are:

1. **Plagiarism Index** – this was calculated as the degree of lexical and semantic matches or overlaps detected by the model for the documents being compared within the score range of 0 and 100.

2. **Precision and Recall**
Accuracy of the model in distinguishing between plagiarized and non-plagiarized contents of the submitted documents was measured with precision and recall metrics. These metrics revealed the quantity of document items flagged as plagiarized that are actually plagiarised in the true sense.

$$precision = \frac{TP}{TP+FP} \quad \text{..... Equ 1}$$

That is, of all flagged plagiarism, how much was actually plagiarized?

$$Recall = \frac{TP}{TP+FN} \quad \text{..... Equ 2}$$

That is, of all actual plagiarism, how much was detected?

These were computed using a confusion matrix, based on the binary classification.

Where:

TP - True Positive; FP - False Positive; FN – False Negative; TN – True Negative;
And in the context of this study:
TP - Plagiarized and correctly flagged; FP - Non-plagiarized but wrongly flagged FN - Plagiarized and not flagged; TN - Non-plagiarized and correctly not flagged

3. F1 Score

To balance precision and recall, the F1-score was used:

$$\text{F1 score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots Equ\ 3$$

4. Accuracy

Accuracy gives an overall measure of how many predictions (both plagiarized and non-plagiarized) were correct.

5. Semantic Similarity Score

The model computed concept overlap between source and target text, especially for light or paraphrased plagiarism. The system was also tested using with primary data - real-world student assignments.

Results and Discussion

Performance of the model is discussed below.

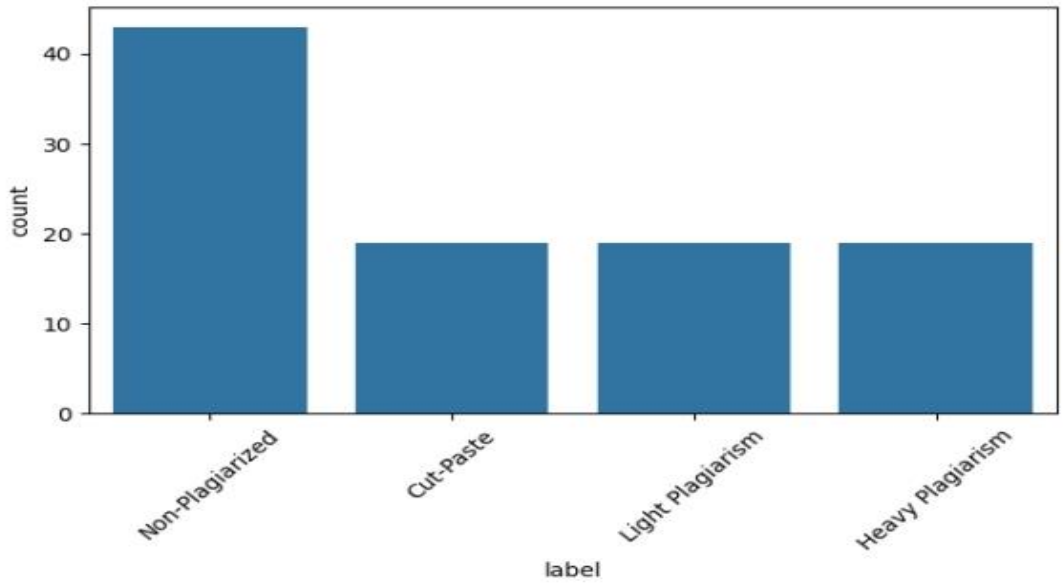

Figure 2: Bar chart showing plagiarism results for ACOPSA dataset

Model Performance Summary

The performances of the Random Forest, XGBoost, and KNN models were compared using accuracy, precision, and F1 score. The results are presented in Table 2. Random Forest (Accuracy: 0.86 , Precision: 0.86, and F1 Score: 0.85) achieved the best overall performance, followed

by XGBoost (Accuracy: 0.80, Precision: 0.82, and F1 Score: 0.80) while KNN yielded the lowest performance (Accuracy: 0.57, Precision: 0.59, and F1 Score: 0.55) and was therefore considered suboptimal.

Table 2 Comparison of Machine Learning Models for Plagiarism Detection

	Model	Accuracy	Precision	F1 Score
1	Random Forest	0.86	0.86	0.85
2	KNN	0.57	0.59	0.54
3	XGBoost	0.80	0.82	0.80

Confusion Matrix Analysis

To further evaluate the model's classification effectiveness across different types of plagiarism, confusion matrices were generated. These matrices help visualize the distribution of correct and incorrect predictions for each plagiarism category: Cut-Paste, Heavy Plagiarism, Light Plagiarism, and Non-Plagiarized content.

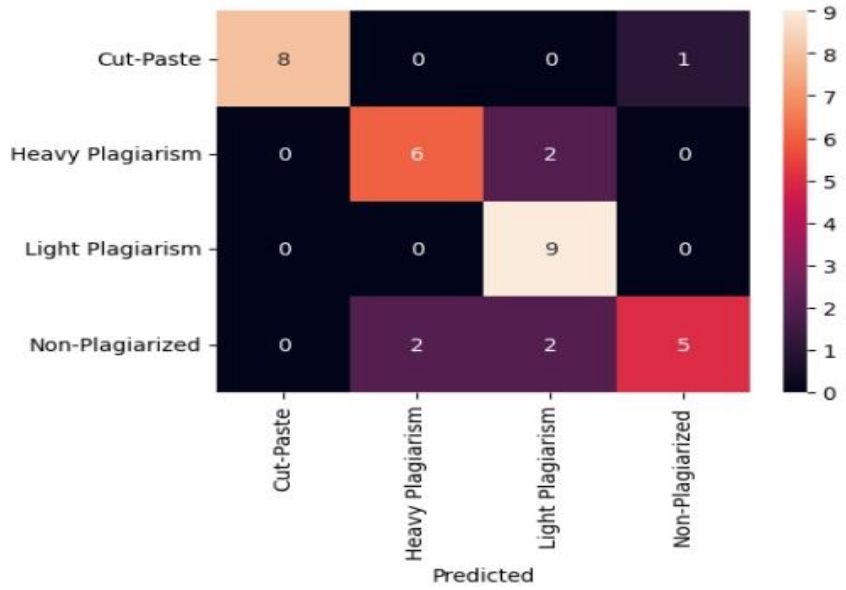

Figure 3: Confusion Matrix for XGBoost Classifier

The confusion matrix for XGBoost shows strong performance across most categories: Cut-Paste Plagiarism was correctly identified 8 out of 9 times, with only 1 misclassified as *Non-Plagiarized*. Heavy Plagiarism saw 6 correct predictions, but 2 were misclassified as *Light Plagiarism*. Light Plagiarism achieved perfect classification with 9 out of 9 correct predictions. Non-Plagiarized content had a moderate misclassification rate, with 2 cases wrongly predicted as *Heavy Plagiarism* and 2 as *Light Plagiarism*. XGBoost demonstrates high sensitivity to *Light Plagiarism* and *Cut-Paste* categories, but shows some difficulty distinguishing *Non-Plagiarized* from *semantically similar* plagiarized content, possibly due to paraphrasing.

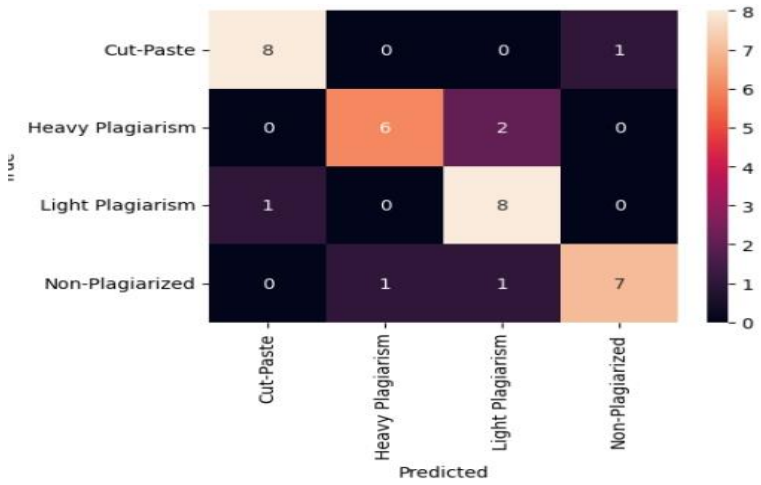

Figure 4: Confusion Matrix for Random Forest Classifier

Cut-Paste Plagiarism was correctly classified in 8 out of 9 cases. Heavy Plagiarism had 6 correct and 2 misclassified as *Light Plagiarism*. Light Plagiarism saw 8 correct predictions, with 1 false positive as *Cut-Paste*. Non-Plagiarized was mostly well identified (7 correct), with only 1 misclassified each as *Heavy* and *Light Plagiarism*. Random Forest shows a balanced performance across categories, particularly in handling *Non-Plagiarized* submissions. However, slight confusion between *Light* and *Cut-Paste* categories suggests limitations in lexical-level differentiation.

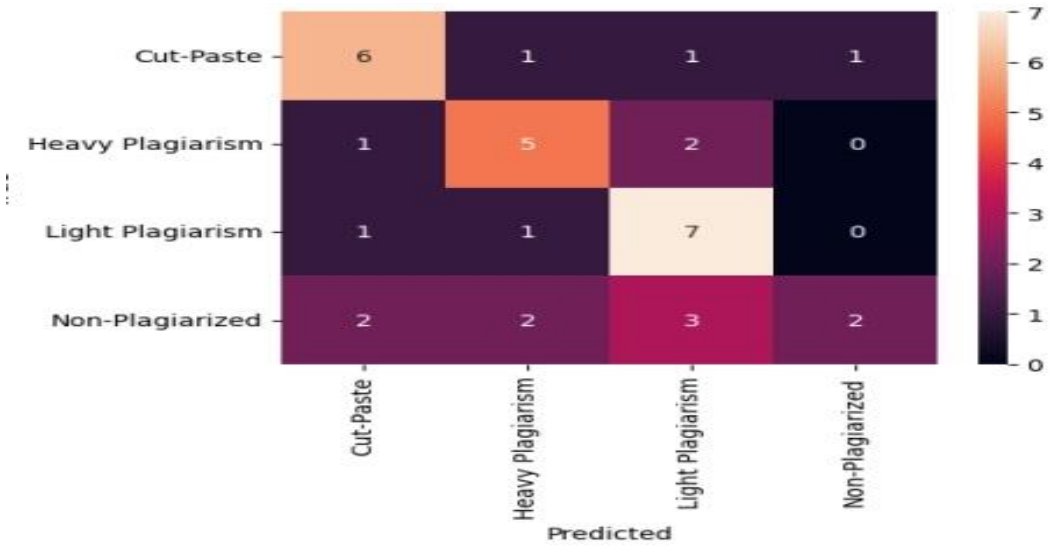

Figure 5: Confusion Matrix for K-Nearest Neighbors (KNN) Classifier

Cut-Paste Plagiarism was detected correctly in 6 cases; 3 were misclassified. Heavy Plagiarism had 5 correct classifications, with 1 predicted as *Cut-Paste* and 2 as *Light*. Light Plagiarism saw 7 correct predictions, with 2 misclassified. Non-Plagiarized content had the highest misclassification rate, with only 2 correct predictions. The remaining were misclassified into *Cut-Paste*, *Heavy*, or *Light* categories. KNN underperformed compared to the

other models. It had difficulty generalizing, especially for *Non-Plagiarized* content, likely due to its reliance on local similarity without deeper contextual understanding.

Table 3 Summary of Confusion Matrix

Model	Strengths	Weaknesses
XGBoost	Excellent at detecting *Light Plagiarism*	Misclassifies *Non-Plagiarized* as *Plagiarized*
Random Forest	Balanced detection, especially *Non-Plagiarized*	Occasional confusion between *Light* and *Cut-Paste*
KNN	Decent for *Light Plagiarism*	Poor differentiation for *Non-Plagiarized* cases

These findings align with the performance metrics in Table 3, further confirming the superiority of Random Forest and XGBoost in detecting various plagiarism categories, with XGBoost showing slightly better precision and F1 score, while Random Forest excelled in handling real non-plagiarized content.

Receiver Operating Characteristic (ROC) Curve

Performances of Random Forest, XGBoost, and K-Nearest Neighbours (KNN) machine learning classification models in detecting various forms of plagiarism using the ACOPSA dataset were also evaluated using ROC (Receiver Operating Characteristic) curve analysis (figure 6), to illustrate how well each model could distinguish among the four forms of plagiarism under consideration, namely: Non-Plagiarised, Light Plagiarism, Heavy Plagiarism, and Cut-and-Paste. The area under this curve enables us to determine the trade-off between the true positive rate and the false positive rate across different value levels. A higher Area Under the Curve (AUC) indicates better model performance.

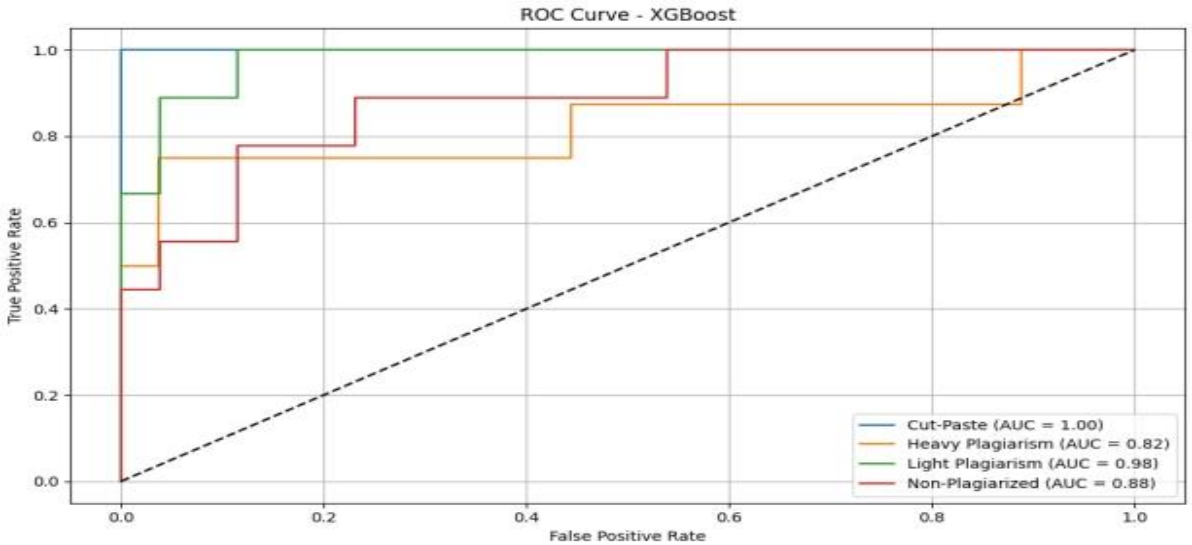

Figure 6: ROC Curve for XGBoost Model

XGBoost demonstrated exceptionally high detection accuracy for *Cut-and-Paste* plagiarism with an AUC of 1.00, indicating perfect classification. It also

performed very well in detecting *Light Plagiarism* (AUC = 0.98), and moderately well for *Non-Plagiarised* (AUC = 0.88) and *Heavy Plagiarism* (AUC = 0.82). See figure 6.

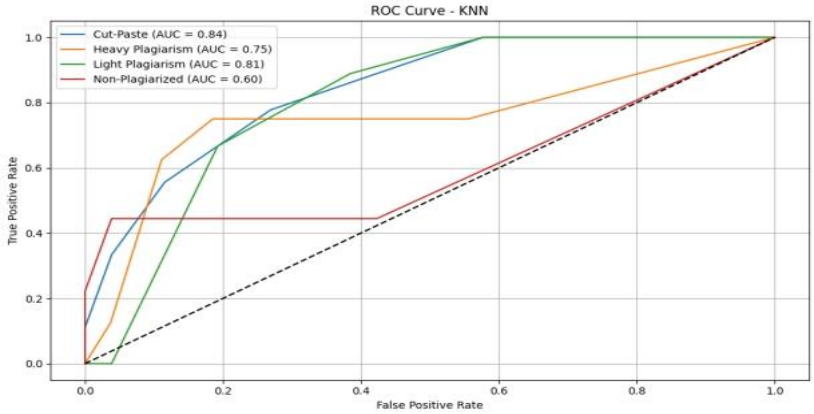

Figure 7: ROC Curve for K-Nearest Neighbour (KNN) Model

KNN underperformed in most categories, especially in identifying *Non-Plagiarised* content, with a low AUC of 0.60, indicating a tendency to falsely label original work as plagiarised. Its AUC values for *Cut-and-Paste* (0.84), *Light Plagiarism* (0.81), and *Heavy Plagiarism* (0.75) were moderate but not reliable enough for academic detection systems (see figure 7).

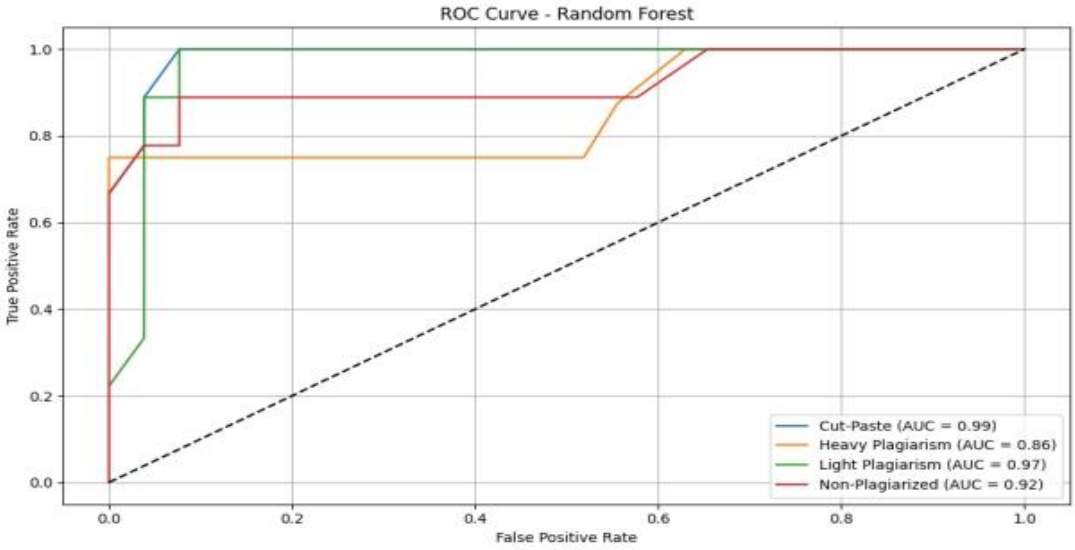

Figure 8: ROC Curve for Random Forest Model

Random Forest, by contrast, offered the most balanced and consistent performance across all categories. It achieved very high AUC values: *Cut-and-Paste* (0.99), *Light Plagiarism* (0.97), *Non-Plagiarised* (0.92), and *Heavy Plagiarism* (0.86). These results confirm the model's robustness and its superior capacity to discriminate accurately between original and

plagiarised text in all forms (see figure 8).

The confusion matrix analyses reveal that while Random Forest achieves balanced performance across all categories, challenges remain in differentiating closely related plagiarism types—especially Light versus Heavy or Cut-Paste plagiarism. Similarly, the need for further refinement to minimize false positives, which is crucial to avoid unfairly penalizing genuine student work is what the moderate misclassification rates for non-plagiarized content highlights.

As opined by Oravec (2023), this study advocates constitution and utilization of institutional policies and educational initiatives aimed at promoting ethical use of AI given the rising challenges of identifying AI generated contents. This requires that detection models are continuously updated and more sophisticated lexical, semantic and cross-language detection features incorporated to adequately combat dynamically emerging forms of plagiarism.

As highlighted in this study, it is imperative that each institution prioritises digitizing and indexing their local contents in their repositories to make them available online to facilitate performance of plagiarism detection tools and increase their detection reach.

Conclusion

This study developed a model that carried out performance analysis and comparison of three machine learning classification algorithms in detecting plagiarism in assignments for online submissions. Random Forest and XGBoost showed strong detection performance, as Random Forest provided the most balanced and reliable detection across different types of plagiarism. The system was particularly effective at identifying clear cases of cut-and-paste plagiarism as well as subtle paraphrasing.

While the AI models demonstrated promising accuracy and precision, some challenges remain, especially in distinguishing between closely related plagiarism categories and avoiding false positives on original work. These limitations highlight the need for continued improvements in model design and access to comprehensive datasets, including local academic content. Future work should consider working with larger datasets, incorporating cross-language semantic similarity detection and institutional advocacy for ethical use of AI in academics and research.

Recommendations

Institutions and researchers are encouraged to regularly review and benchmark existing plagiarism detection tools to identify their limitations and uncover opportunities for enhancement with AI models. Future development efforts should focus on designing hybrid AI models that integrate both lexical and semantic similarity techniques, which will improve plagiarism detection of paraphrased and cleverly disguised contents even those involving cross-language academic dishonesty Additionally, it is important for developers to simulate and validate the performance of these models using diverse and realistic datasets, including publicly available datasets such as ACOPSA as well as local student submissions.

References

Ahmad, R., Odafe, E., & Zhang, Y. (2024). *Efficient semantic matching of academic documents using vector representations*. *Journal of Computational Linguistics in Education*, 18(1), 33–49.

AmberBlog (2025) Exploring the evolution of plagiarism detection tools https://amberstudent.com/blog/post/evolution-of-plagiarism-detecting-tools#:~:text=Plagiarism%20checkers%20evolved%20to%20use,matches%20in%20its%20extensive%20database. Last visited on 2/11/25.

Chen, L., & Raji, S. (2023). *From string match to meaning match: Advances in semantic plagiarism detection. AI & Ethics in Academia*, 5(2), 112–127. https://doi.org/10.1016/aea.2023.112

Dahmen, J., Kayaalp, M., Ollivier, M., Pareek, A., Hirschmann, M. T., Karlsson, J., & Winkler, P.

W. (2023). Artificial intelligence bot ChatGPT in medical research: The potential game changer as a double-edged sword. *Knee Surgery, Sports Traumatology, Arthroscopy, 31*, 1187–1189.

Foltənek, T., Dlabolov6, D., Anohina-Naumeca, A. *et al* (2020). Testing of support tools for

plagiarism detection. *Int J Educ Technol High Educ 17(46).* https://doi.org/10.1186/s41239-020-00192-4

Gillard, C., & Rorabaugh, P. (2023). You're not going to like how colleges respond to ChatGPT. *Slate*. Retrieved February 24, 2025, from https://slate.com/technology/2023/02/chat-gpt-cheating-college-ai-detection.html

Heinrich, P., & Maurer, H. (2000). *Case studies on semantic equivalence and plagiarism detection. Journal of Universal Computer Science*, 6(4), 412–427.

Huang, K. (2023). Alarmed by AI chatbots, universities start revamping how they teach. *New York Times*. Retrieved February 24, 2025, from https://www.nytimes.com/2023/01/16/technology/chatgpt-artificialintelligence-universities.html

Ibrahim, K. (2023). Using AI-based detectors to control AI-assisted plagiarism in ESL writing:

"The Terminator Versus the Machines". *Lang Test Asia 13 (46).* https://doi.org/10.1186/s40468-023-00260-2

Kumar, A. (2021). The role of AI in plagiarized text. Learning Hub. Retrieved February 26, 2025, from https://learn.g2.com/ai-for-plagiarism

Oravec, J. A. (2023). Artificial intelligence implications for academic cheating: Expanding the

dimensions of responsible human-AI collaboration with ChatGPT and Bard. Journal of Interactive Learning Research, 34(2), 213–237.

Nakpih C. I. (2024) A modified vector space model for semantic information retrieval. NLP Journal 8(5).

Singh, R., & Kumar, T. (2022). Limitations of lexical matching in plagiarism detection: A comparative evaluation. International Journal of Digital Education, 15(3), 88–102.

Sozon, M., Sia, B.C., Pok, W.F. and Alkharabsheh, O.H.M. (2024), Academic integrity violations in higher education: a systematic literature review from 2013–2023, Journal of Applied Research in Higher Education *, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/JARHE-12-2023-0559*

Tariq, A., Zhou, L., & Obi, N. (2022). Semantic similarity detection using contextual embeddings

for paraphrased plagiarism. Transactions in Knowledge and Data Engineering, 34(9), 1901–1913.

Turney P.D. & Pantel P. (2025) From Frequency to Meaning: Vector Space Models of Semantics. *Journal of Artificial Intelligence Research 37(1)*

Vani, K and Gupta, D. (2016) Study on Extrinsic Text Plagiarism Detection Techniques and Tools. Journal of Engineering Science and Technology Review 9 (4) 150 – 164

Yilmaz S., & Fernбndez F. (2024). Evaluation of Turkish academic and student attitudes on plagiarism: validity and reliability of the plagiarism attitude scale. *Journal of empirical research on human research ethics 19(1-2)*

Zhao, C., & Mensah, J. (2025). Ontology-aided semantic plagiarism detection in domain-specific academic writing. Educational Technology Advances, 11(2), 75–91.

Zihang Z., Boqing G., & Liqiang W. (2025**).** Attention to Neural Plagiarism: Diffusion Models Can

Plagiarize Your Copyrighted Images. *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2025, pp. 19546-19556*

.