DESIGN AND IMPLEMENTATION

OF

AUTOMATIC GRADING SYSTEM

BY

ONIGBODE NATHAN

BHU/21/04/05/0027

B. SC. COMPUTER SCIENCE

NOVEMBER, 2024

PROJECT RESEARCH PROPOSAL

ON THE

DESIGN AND IMPLEMENTATION OF AN AUTOMATIC GRADING SYSTEM

BY

ONIGBODE NATHAN
BHU/21/04/05/0027

A PROJECT SUBMITTED TO THE DEPARTMENT OF COMPUTER SCIENCE, FACULTY OF COMPUTING, BINGHAM UNIVERSITY, KARU, NASSARAWA STATE, NIGERIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF BACHELOR DEGREE IN COMPUTER SCIENCE

DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF COMPUTING
BINGHAM UNIVERSITY, KARU
NIGERIA

NOVEMBER, 2024

DECLARATION

I, **ONIGBODE NATHAN**, with Registration number **BHU/21/04/05/0027** hereby declare that the content of this project report has been carried out by me in the Department of Computer Science, Bingham University Karu, under the supervision of MRS. OLUWATOYIN ADEYEMO. The information derived from the literature has been duly acknowledged in the text and a list of references provided. No part of this project was previously presented for another degree or diploma at this or any other institution.

……. …….

ONIGBODE NATHAN **Date**

BHU/21/04/05/0027

CERTIFICATION

This is to certify that this project report written by **ONIGBODE NATHAN** with Registration Number **BHU/21/04/05/0027** titled "**Automatic Grading System**" meets the requirements governing the award of the degree of Bachelor of Science in Computer Science and is approved for its contribution to knowledge.

_____ _____

MRS. OLUWATOYIN ADEYEMO **Date**
SUPERVISOR

_____ _____

DR. IBRAHIM YAKUBU **Date**
LEVEL COORDINATOR

_____ _____

DR. ADAMU S. USMAN **Date**
HEAD OF DEPARTMENT

ACKNOWLEDGMENT

I would like to express my sincere gratitude to God Almighty for the gift of life. I wish to express my sincere appreciation my project supervisor Mrs. OLUWATOYIN ADELAKUN ADEYEMO, for the guidance, inspiration and continuous support during the period of my project. I would like to thank the Head of Department of Computer Science Dr. Adamu S. Usman. My profound gratitude goes to the Departmental Level Coordinator Dr. Yakubu Aliyu Ibrahim and all the lecturers in the Department which include Dr. (Ms.) Anna B. Hassan, Dr. Faki A. Silas, Dr. Onu Egena. Dr. Oluwasegun Ishaya Adelaye, Dr. Victor Kulumph, Mrs. Oluwatoyin Adelakun Adeyemo, Mr. Musa Yusuf, Mr. Maikori Ezekiel Jenome, Mr. Barka T. Fori, Mr. Ibrahim Lawal and Mrs. Ipole Nancy. I also thank the Technologists in the Department Mr. Joseph Oladele Aremu, Mr. Ngale Langthong and Mr. Sharack Akoh. I also appreciate the Departmental Secretary Mrs. Talatu Bowman. A big thank you to all the teaching and non-teaching staff of the Faculty of Science and Technology Bingham University. I would like to thank my loving parents Mr. and Mrs. ONIGBODE, for giving their support financially, morally, emotionally and spiritually. Also say thank you to my beloved sibling Naomi and Faith for being there for me always. I thank my fellow Students from the Department of Computer Science for their support and logical contributions in the various parts of my project. Finally, I appreciate everyone whose name were not mention but has in one way or the other contributed to the success of this project report. May Almighty God bless them all.

ABSTRACT

This technical report is dedicated to my loving parents, whose unwavering support, encouragement, and belief in my abilities have been a constant source of motivation throughout my academic journey.

All thanks belong to God Almighty the most merciful, the beneficent. I wish to express my profound gratitude to Almighty God for his assistance throughout my academic line.

My special appreciation goes to the school managements, my lecturers, and all the project supervisors for their efforts.

I also dedicate this project to my sisters Naomi and Faith, and to my friends and relatives.

TABLE OF CONTENTS

.

TABLE OF FIGURES

No table of figures entries found.

CHAPTER ONE

INTRODUCTION

1.1. Background of Study

The general area of study of my project is focused on the use of machine learning in the educational sector. While education plays a vital role in societal development, with assessments being a key aspect of evaluating learning outcomes, traditional assessment methods, particularly those involving essay or short-answer responses, often require substantial manual effort from educators (Singh et al., 2020). These methods can be time-consuming, subjective, mentally stressful and prone to human error (Sharma & Patel, 2019). To address these challenges, automated grading systems have emerged as a promising solution, offering efficiency, consistency, and scalability in assessing responses of students (Chen & Xie, 2021).

In recent years, advancements in natural language processing (NLP) and machine learning (ML) have enabled the development of sophisticated systems capable of interpreting, understanding and evaluating text content (Mishra et al., 2022). Automated grading systems leverage these technologies to assess student responses by analyzing linguistic patterns, processing text elements, content relevance, and structure, thereby reducing the burden on educators (Lee et al., 2020).

A critical aspect of any grading system is to ensure academic integrity. Plagiarism, the act of using someone else's work without proper acknowledgment, undermines the fairness of evaluations and the credibility of education systems (Park, 2003). The integration of plagiarism detection mechanisms into automated grading systems can address this issue by identifying copyright content and providing detailed reports (Das et al., 2019). Plagiarism detection tools typically use techniques such as text similarity

analysis, semantic matching, and large data content of academic material to detect potential infringements (Zhao et al., 2021).

Numerous studies have highlighted the advantages of automated grading systems. These include faster processing times, the ability to handle large volumes of responses, and impartiality in scoring (Singh et al., 2020). Despite these benefits, accurately understanding creative answers and addressing concerns regarding over-reliance on technology in assessments are some challenges that remain (Sharma & Patel, 2019).

The increasing use of online learning platforms and digital education tools has further emphasized the need for reliable automated grading systems. In this context, a system that combines grading with an inbuilt plagiarism checker becomes an essential tool (Mishra et al., 2022). Such a system not only evaluates the quality and accuracy of responses but also promotes originality and adherence to ethical standards in academic work (Chen & Xie, 2021).

The present study aims to design and develop an Automated Grading System with an inbuilt Plagiarism Detector, leveraging state-of-the-art technologies to ensure efficient and fair assessment. By doing so, it will improve the quality of education, support educators, and foster a culture of integrity among learners.

1.2. Statement of Problem

The problem results from the work load of students' response assessment on lecturers and how grading assignments can reduce the effective time of the lecturer while maintaining the students' results authenticity. Due to the stress and time spent on grading assignments (especially for a large group of students) the lecturer's declining stamina and attention will reduce the efficiency of further grading, leading to incorrect assessment of the responses. Another perspective is from the students who, has a result, get inaccurately distributed marks.

This problem begins to occur when the lecturer in question doesn't have enough of a resource (time, patience, attention span, mental stability) to begin or continue grading his or her students' responses to an assignment. This problem affects the lecturers and the students which is the educational sector at large. To resolve this problem, an artificial intelligence which has been trained on natural language processing and deep learning models for text analysis will be used to develop a system that can grade short answers and essays automatically.

1.3. Aim and Objectives

The aim of the Automated Grading System is to ease the burden of grading on lecturers by using machine learning to assess student responses (especially for open-ended, short-answer, and essay-style questions). This approach would provide insightful and faster feedback, particularly for subjective answers, beyond what traditional grading offers.

The objectives to be carried out to ensure the success of this project includes:

i. **Collect** appropriate data to ensure the accuracy and efficiency of the model.
ii. **Design** a natural language processing model capable of recognizing everything from typographical errors to complex phrases.
iii. **Implement** the model using various machine learning frameworks and tools.
iv. **Test** the model with a variety of essays of different complexities and topics.
v. **Evaluate** the effectiveness of the model after the tests have been concluded.

1.4. Significance of the Study

The study is important to the society and has several impactful benefits for education. It ensures the efficiency in the sense that lecturers can reduce the unnecessary use of high-quality resources, saving significant time on lesson planning and evaluation. It will also improve the quality of education because

when the students are aware of an anti-plagiarism tool, it will result in more personalized responses, these personalized responses will help students engage with material thereby improving their receptiveness to learning by their pace and style. From one perspective, the model will ensure scalability by automating categorization and recommendations in turn, supporting large and diverse repositories of responses.

1.5. Scope of Study

The study will use the web application "Google Classroom" as the scope and case study. The project will be centered on how to extract students' responses from the assignment deck and analyze them immediately the time for submission has elapsed. From being just a stand-alone application, it will expand to being an extension on the web browser so that it will automatically begin operation when connected with the Google Classroom API. On startup, the project will focus on singular submissions before extending to batch response collection. The project will be centered around the lecturers and students featuring a friendly user interface with exceptional reporting functionality to generate insights and correction where needed. Unfortunately, due to complexity, the model will not be compatible for image processing. This will be the major limitation of the model as it will be very difficult to train the model to recognize images and grade them.

1.6. Organization of the Study

CHAPTER ONE – INTRODUCTION: The study is about the use of natural language processing and machine learning to create an artificial intelligence model that will automatically grade students' responses. The background of study highlights the role of education and the effect of evaluation in societal development. The statement of problem centers on how the workload on lecturers when grading tasks and providing feedback the traditional way can reduce the effectiveness of that lecturer in the long run. The study aims at easing that burden by automating the process while providing faster and more insightful feedback. The study is significant in the sense that it will improve the quality of education while reducing the human resources needed for said improvement. Initially the study will begin as a

stand-alone application which will only process single responses at a time but will scale up to a web extension that will grade thousands of submissions at an instant.

CHAPTER TWO – LITERATURE REVIEW: The research data is mostly going to be collected from datasets in the online website "Kaggle.com" and previous models from the same website will be studied and used as preliminary thesis for the development of the system. A literature review will be conducted to explore existing automated grading and plagiarism detection systems. Suitable NLP and ML frameworks for system development will be identified and datasets for training the grading and plagiarism detection algorithms will be collected and prepared resulting to the design of a prototype for testing and refinement. Text Preprocessing, Tokenization**,** Embedding, Representation Learning, Sentence and Document-Level Analysis, Similarity and Relevance Analysis**,** Sentiment and Grammar Analysis are key words that will be used in the study. The development of this system is supported by research on NLP and ML models that can process and evaluate natural language efficiently, the increasing availability of large datasets, enabling effective training of algorithms to identify semantic similarity and plagiarism and finally a growing demand for scalable assessment solutions in online education platforms.

CHAPTER THREE – SYSTEM DESIGN AND METHODOLOGY: Quantitative methods will be implemented during the research phase and all data that will be used will be grounded on secondary (pre-existing) datasets. The system will first clean the input text and then break it down to tokens, which will be transformed into vector representation that captures semantic relationships allowing context and meaning understanding. Transformers will be implored to allow larger chunk processing which makes it suitable for essay grading. Semantic similarity scoring will allow the system to measure how the student's answer aligns with a models answer finally a feedback system will be developed to enable adaptability of the model. Also the model will integrate the use of information security risk management frameworks such as ISO 27001 to ensure data privacy and security by protecting the student's data during grading

and plagiarism checks. After the implementation of all these concepts, an experimental design will be formulated.

CHAPTER FOUR – IMPLEMENTATION AND TESTING: This chapter presents the output of the model. It is majorly dependent on the design and methodology stage where all the work done is fully explained from the application to the results of simulations done and even the statistical approaches used up until its modelling. The acceptance, strength and effect of proposed output is presented here and is fully your work so no reference except for evaluation and discussion purpose. It is made up of screenshots, figures, tables, relevant illustrations, data presentation and analysis. This chapter presents the research findings.

CHAPTER FIVE - SUMMARY, CONCLUSION AND RECOMMENDATION: The summary should present highlights of the major findings while the conclusions give an inference drawn from the findings. Challenges encountered during the study should be indicated. Conclusions should be drawn on the basis of the data presented and analyzed, and policy. Recommendations should be based on the major findings of the study and stated in precise terms. It should list possible ways of solving problems identified as well as highlight areas for further research.

CHAPTER TWO

LITERATURE REVIEW

2.1. Introduction

The integration of automated grading systems (AGS) with inbuilt plagiarism detection has become a critical area of research in educational technology. As digital learning platforms expand, there is an increasing need for efficient, scalable, and reliable tools to assess student work while ensuring academic integrity. AGS leverages on technology to evaluate student submissions, such as essays, code, or assignments. The systems aim to reduce the manual effort required by educators while ensuring consistent and timely feedback. AGS often utilize machine learning, natural language processing and artificial intelligence to analyze submissions. As educational institutions increasingly adopt online learning platforms, the demand for robust automated grading tools has grown significantly. However, ongoing research and innovation are essential to overcome challenges and ensure these systems remain robust, ethical, and adaptable in diverse educational contexts.

2.2. Historical Context of the Research Topic

The concept of automated grading emerged in the mid-20th century with early efforts to create machine-readable answer sheets for multiple-choice questions. Over time, technological advancements enabled the evaluation of complex student work, such as essays and coding assignments. Pioneering systems like Page and Peterson's Project Essay Grade (PEG) in the 1960s laid the groundwork for modern automated essay scoring. Recent studies focus on improving the interpretability and accuracy of AGS. Research by Phandi et al. (2015) explored deep learning models that outperform traditional automated essay scoring systems in contextual understanding. Similarly, hybrid approaches combining human and machine grading are gaining traction to address limitations.

Plagiarism detection, on the other hand, gained prominence in the 1990s with the rise of digital content. Tools like Turnitin revolutionized how academic institutions addressed plagiarism by comparing submissions against vast databases of texts. The convergence of AGS and plagiarism detection began in the early 2000s, driven by the growing need for holistic assessment tools.

2.3. Key Concepts and Definitions

1. Automated Grading Systems (AGS): Software systems that use algorithms and artificial intelligence to evaluate and score student submissions.

2. Plagiarism Detection: Techniques used to identify similarities between a submitted work and existing content to ensure originality.

3. Natural Language Processing (NLP): A branch of AI that focuses on the interaction between computers and human language, essential for analyzing textual submissions in AGS. The development of this system is supported by research on ML models that can process and evaluate natural language efficiently,

4. Academic Integrity: Adherence to ethical standards in academic work, including the avoidance of plagiarism. These systems address the growing demand for efficient, scalable, and fair assessment tools.

2.4. Review of Related Literature

Automated Grading Systems: Research by Shermis and Burstein (2013) demonstrated that AES tools like E-Rater use NLP to evaluate essays with accuracy comparable to human graders. Other tools, such as Gradescope, extend this functionality to diverse assignment types, including programming tasks.

Plagiarism Detection: Maurer et al. (2006) highlighted the evolution of plagiarism detection algorithms, from basic string matching to advanced AI techniques for identifying paraphrased content. Tools like Grammarly and PlagScan integrate these methods to ensure comprehensive detection.

Integration of Grading and Plagiarism Detection: Studies emphasize the importance of combining these functionalities. For instance, hybrid systems like WriteCheck evaluate assignments while detecting unoriginal content, streamlining the assessment process.

2.5. Methodology

Research methodologies in this field include:

i. Quantitative Analysis: Evaluating the performance of AGS and plagiarism detection tools using metrics like precision, recall, and accuracy. Quantitative methods will be implemented and all data that will be used will be grounded on secondary (pre-existing) datasets.

ii. Qualitative Analysis: Assessing user satisfaction and system effectiveness through educator and student feedback.

iii. Comparative Studies: Comparing automated systems with human evaluation to identify strengths and limitations.

iv. Case Studies: Documenting the adoption of integrated AGS tools in educational institutions to understand their impact.

2.6. Empirical Review

Empirical studies reveal that automated systems significantly reduce grading time while maintaining accuracy. For instance, Phandi et al. (2015) demonstrated that Bayesian frameworks outperform traditional scoring models in understanding essay context. Additionally, plagiarism detection tools like Turnitin report detection rates exceeding 90% for copied content. However, challenges such as false positives and contextual misinterpretations persist. A study by Maurer et al. (2006) emphasized the importance of integrating plagiarism detection in AGS, noting its effectiveness in curbing unethical practices. Modern systems extend their capabilities by using AI to paraphrase detection, which captures subtle forms of plagiarism.

Research by Shermis and Burstein (2013) highlighted that automated essay scoring (AES) systems like E-Rater and IntelliMetric utilize NLP models to assess essays with performance comparable to human graders. Inbuilt plagiarism detection is a critical feature of AGS. Advanced plagiarism detection systems, such as Turnitin, PlagScan, and Grammarly, employ algorithms like shingling, hashing, and sequence matching to detect overlaps in text.

For non-textual assignments, such as coding tasks, automated graders like AutoGrader or Gradescope use static and dynamic analysis to evaluate code quality, correctness, and efficiency. Emerging deep learning techniques further enhance grading by identifying patterns and anomalies in student responses.

2.7. Conceptual Review

The conceptual framework of integrated AGS systems is rooted in AI and educational psychology. The system aims to replicate human grading by evaluating content quality, coherence, and originality. Key elements include:

Grading Algorithms: Designed to assess specific criteria, such as grammar, content relevance, and formatting. The system will clean the input text and break it down to tokens, which will be transformed into vector representation that captures semantic relationships allowing context and meaning

understanding. Transformers will be implored to allow larger chunk processing, making it suitable for essay grading.

Similarity Detection Algorithms: Techniques like fingerprinting and semantic analysis used for plagiarism detection. Semantic similarity scoring will allow the system to measure how the student's answer aligns with a models answer.

User Interfaces: Intuitive platforms that enable educators to review and override automated decisions when necessary and a feedback system will be developed to enable adaptability of the model.

Information Security Risk Management Frameworks: Also the model will integrate the use of ISO 27001 to ensure data privacy and security by protecting the student's data during grading and plagiarism checks.

2.8. Debates and Controversies

1. Accuracy and Bias: Critics argue that AGS systems may reinforce existing biases in training data, leading to unfair evaluations as it may misinterpret nuanced arguments, cultural contexts, or creative writing styles.

2. Ethical Concerns: The use of plagiarism detection tools raises privacy issues, as submissions are often stored in external databases.

3. Reliance on Automation: Some educators believe that overdependence on AGS may undermine the role of educators in providing personalized feedback.

4. False Positives in Plagiarism Detection: Common phrases or shared knowledge often trigger plagiarism alerts, leading to unnecessary disputes. Plagiarism detection tools may also occasionally flag original content as plagiarized due to common phrases or technical jargon.

2.9. Practical Implication

The integration of AGS with plagiarism detection has far-reaching implications for education:

i. Efficiency: Institutions can manage large-scale assessments with minimal human intervention. Automated grading significantly reduces the time educators spend on grading, especially in large-scale assessments.

ii. Scalability: AGS can handle large volumes of submissions, making them ideal for Massive Open Online Courses (MOOCs) and for online learning platforms.

iii. Standardization: AGS ensures consistent grading across diverse student groups. The combination of automated grading with plagiarism detection streamlines the evaluation process.

iv. Integrity Assurance: Built-in plagiarism detection fosters a culture of originality and ethical conduct among students. This will uphold academic standards by identifying and addressing copied content.

v. Consistency: AI-based grading ensures objective evaluation, minimizing human bias which represents a transformative approach to academic assessment.

REFERENCES

Chen, X., & Xie, Y. (2021). Advancements in automated grading systems using NLP and machine learning: A review. *Journal of Educational Technology, 45*(2), 124–138. https://doi.org/10.1016/j.jedutech.2021.05.003

Das, A., Kumar, R., & Gupta, M. (2019). Plagiarism detection in digital academic environments. *International Journal of Academic Integrity, 6*(3), 45–57. https://doi.org/10.1080/ijai.2019.1004

Lee, H., Park, J., & Choi, S. (2020). Machine learning models for automated essay scoring and their limitations. *Computers in Education, 58*(4), 78–89. https://doi.org/10.1016/j.compedu.2020.101501

Mishra, S., Verma, P., & Rao, A. (2022). Integrating automated grading with plagiarism detection: Challenges and prospects. *Computers and Education Review, 12*(2), 57–73. https://doi.org/10.1016/j.cer.2022.04.002

Park, C. (2003). In other (people's) words: Plagiarism by university students—Literature and lessons. *Assessment & Evaluation in Higher Education, 28*(5), 471–488. https://doi.org/10.1080/0260293032000120352

Sharma, R., & Patel, S. (2019). The impact of AI-based grading systems on education. *Journal of Educational Research and Development, 39*(7), 112–128. https://doi.org/10.1016/j.jerd.2019.03.009

Singh, V., Kumar, S., & Roy, T. (2020). Automated grading systems: An overview of trends and challenges. *International Journal of Advanced Education Research, 4*(6), 99–118. https://doi.org/10.3126/ijedres.2020.0466

Zhao, L., Wang, Y., & Han, J. (2021). Semantic analysis in plagiarism detection: Advances and applications. *Journal of Academic Ethics, 19*(3), 215–232.

Shermis, M. D., & Burstein, J. (2013). *Handbook of automated essay evaluation: Current applications and new directions.*

Maurer, H. A., Kappe, F., & Zaka, B. (2006). Plagiarism—a survey. *Journal of Universal Computer Science*, *12*(8), 1050-1084.

Phandi, P., Chai, K. M. A., & Ng, H. T. (2015). Flexible domain adaptation for automated essay scoring using a Bayesian framework. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing.*