Automated Assessment System Using Machine Learning Libraries

Victor Adebola Omopariola, Chukwudi Nnanna Ogbonna & Felix Uloko

Veritas University, Abuja, NIGERIA
Faculty of Natural and Applied Science

Monday J. Abdullahi

Airforce Institute of Technology, Kaduna, NIGERIA
Department of Computer Science

Abstract

Assessment and the grading of students is a task that has been done for as long as school has existed. This was previously done by teachers in primary and secondary, lecturers for institutions like JAMB and lecturers in schools. Up until now students' marks were influenced by other external factors such as bad handwriting, lengthy paragraphs, roundabout way of speaking rather than going straight to the point and the sheer number of assignments the lecturer has to mark. This has resulted in students getting lower or higher marks than they should be awarded. This project is to create an ML (Machine Learning) powered assessment system that will take the assignment questions and the marking scheme and award the student the marks similar to what the ideal lecturer would have given. This will also reduce the time the lecturers spend on marking and ensure the students get their results on time. This project will be made with Python and machine learning and will be tested with a number of potential answers to the questions and their grading's. This will enable system to be able to grade assignments as soon as they are uploaded. This research will be limited by the fact that the system can only handle the marking of short sentences accurately and not long paragraphs. The system is also limited by the fact that it can only mark with the aid of the marking scheme and not without it so it is not a truly intelligent model in that regard. The research showed that the system is indeed capable of obtaining the similarity between two paragraphed answers provided but it needs extras to produce the most accurate results.

Keywords: assessment, automated assessment system, machine learning library.

1. Introduction

1.1 Background

Assessments generally refer to the tools educational overseers or educators use to evaluate, measure or determine the educational capacity of a student, the readiness of a student to learn what has been taught to him/her over time and the educational needs of the student being taught. According to Stassen et al., assessment has defined as "The systematic collection and analysis of information to improve student learning." As humans' assessments is very important

for tracking progress for all ages. Generally, this has been done through the use of assessments created by various educational bodies and the various standards set. There are different tests that directly correlate to how much a student has learned about the concepts or information the educator is trying to teach them. There are various testing methods which have been developed throughout the years to determine the level of knowledge and creativity of students. Examples of these testing methods include peer-assessments; this is when a student marks another's assessment using a set of rules or guidelines this helps to improve a student's judgement and learn the processes to a result being awarded. Presentations which usually involve the student delivering a piece in front of either a class or assessors and they get graded for it, discussions mostly in the form of a debate to gauge the knowledge of the students especially when their perspective is being challenged, reports and the most common written assessments; these can come in the form of time constrained individual assessments and these have the unintended effect of surface learning and cramming and can come in various forms such as open-book, in-class assessment or take-home assessments. Generally, assessments cover a range of subjects which include; Expressive arts, Health and wellbeing, Languages (including English, French, classical languages and modern languages), Mathematics, Religious and moral education, Sciences, Social studies, Technologies. Assessments are done periodically and at specified intervals set by either the school or assessment body, such time could be at the end of a school year or at the end of a semester but one thing they all have in common is that they occur at key points in a student's learning journey. Assessment is usually split up into two categories or two purposes (these differ from the method the assessments are carried out in), specifically summative assessment and formative assessment. Summative assessments are done at the end of a course and as a result it "sums up" everything the student has learned from the beginning of the course to the end of the course. They are generally done with the use of comprehensive final assignments or papers. The second type is formative assessment which is done during the students learning time this kind of assessment is done in order to enhance the learning experience of the student. This kind of assessments are done for the sole purpose of sharing the results back to the students so they can understand their strengths and weaknesses and reflect on them. This type of assessment typically includes things like coursework and others. Grading is not supposed to be the major aspect of assessments as assessments is supposed to make sure the students attain the knowledge required of the course. Also Grading does not tell you about the students individual learning outcomes that have been achieved. Grades are now one of the most important parts of any schooling although they may not accurately reflect the level of the student's skill or understanding. The job market revolves around how well you did in university or college with top grades such as "First Class" or "Second Class Upper" being preferred across universities regardless of the rank of the university the grade is obtained from. According to Allen et al. (2001), "the larger the variability in grading practices from teacher to teacher and from school to school, the more limited the value grades have as guides for planning the academic and career futures of students."

1.2 *Problem statement*

From the problem statement above, we see hos crucial the grading process is to the future of students and more often than not scripts are graded differently by different teachers in the same discipline. This project aims to increase ease of marking assignments and CA's and provide the students with an indisputable result and also to increase the confidence of the students when writing assignments and the performance of the students over time.

Open Journal for Information Technology, 2023, *6*(2), 97-122.

_____

1.3 *Research questions*

1. How efficiently can a system grade a student's assignment?

2. How accurate is the grading of a student's assignment by an artificial body?

3. The influence of computer assisted grading on the morale of the students and the education system.

4. The time taken to deliver assignments after they have been graded automatically by the system.

1.4 *Research aim and objectives*

- To investigate whether the involvement of a system to mark and grade students provides a true representation of their skill and increases their confidence in writing assignments.
- To determine if a system is capable of marking assignments using current existing algorithms.
- To investigate if a paragraph styled answers can be appropriately compared without error.
- To find out if existing algorithms can compare answers of varying lengths.

1.5 *Research motivation*

The motivation for this research is based on the fact that many factors play a role in the grading of a student's paper such as handwriting, lecturer mood and the fact that assessments can be subjective depending on whether the lecturer is strict or lenient. This all affects a student's final grade and result so through this project I am attempting to solve some of the issues with the assessment process leading to fairer results and increased satisfaction with the schooling system.

1.6 *Significance of the research*

This research can be applied by all various universities, secondary schools and even the ministry of education to remove disparities when the assessment system and increase the confidence of the students in the system.

1.7 *Delimitation of the research*

This research will be limited by the fact that the system can only handle the marking of short sentences accurately and not long paragraphs. The system is also limited by the fact that it can only mark with the aid of the marking scheme and not without it so it is not a truly intelligent model in that regard. The system will be limited by the fact that it can only compare answers in English and not in other languages or disciplines. Another limitation is the fact that the documents have to only contain the answers and no other additions such as a cover page for the system to work effectively.

2. *Literature review*

Assessment is an integral part of the learning process and an accurate method of gauging how much a student has learned throughout the duration of the course. This process has however not been without issues. As I stated in the background of this study, according to Allen et al. (2001),"the larger the variability in grading practices from teacher to teacher and from school to school, the more limited the value grades have as guides for planning the academic and career futures of students." More importantly validity and reliability of grading practices used in the marking process have had a profound effect on the futures of students Allen et al. (2001) also said in their paper that "Since important decisions are often based on a student's grade, unreliable and invalid grades may result in dire consequences for the student. Invalid grades that communicate an understatement of the student's understanding may prevent a student with ability to pursue certain educational or career opportunities." He even goes further to state that there could be consequences of giving a student more marks than he/she deserves this could lead to the student, after graduation being put in situations that his/her grade says they are fit for but in reality they are underprepared or they have an inadequate level of information required for the role or position. Also supporting my theory that the grading system of schools does not take account the status of the school he states "Research indicates that when compared to schools in more affluent areas, students in low SES schools receive grades that are two letter grades better than students in affluent schools when national standardized scores are held constant."

Cizek (1996) in his paper states that one of the crucial aspects of grading that needs to be touched is the training of teachers in grading practices based on sound measurement principles relevant to their classroom lives. In the grading process it is shown that the classroom actually affects a final grade a student gets in addition to assignments. Cross et al. (1996) stated that "Some studies have found that 2 out of 3 teachers believe that effort and student conduct and attitude should influence final grades of students." This goes to show that while a student may have outstanding performances in written assessments if they do not show substantial effort in their classes the lecturers are unlikely to award them the marks they deserve even if they show the adequate amount of knowledge required to earn the grade. Also, regarding the problem that the grading process is subjective. Allen et al. (2001) quote that "All grading is at some level inherently subjective. However, teachers need to recognize the subjective factors in order to reduce them as much as possible to increase the objectivity and validity of their assessment and grading practices." This could be due to problems such as student handwriting, student behavior, the number of scripts given to the teachers to assess and others. Regardless of the method taken to reduce subjectivity it still won't change that fact. This is why the proposal of a system to mark scripts automatically and fairly is an important one. Systems won't be affected by the problems suffered by the lecturer or won't have prior experience with the student outside the questions and answers provided to it.

Though this project aims to remove bias from the marking of assignments it doesn't affect the in-class scores of the students during assignments. For example, if a faculty were awarding 30 marks to classwork's and tests, 40 marks to in class contributions and participation while the rest 30 marks is gotten from the assignments, this will ensure a student gets the best marks from the classwork's and the assignments whereas the other 40 is within the direct control and discretion of the lecturer or teacher. This would also correlate with Allen et al. (2001) study which showed that "This would seem to imply that a grade is used to communicate not only how much content knowledge one has achieved, but also how well one has complied with the teacher's requirements." In my opinion this helps to stress that fact that although assignment grading alone is not enough to give a student an excellent result it should not be overlooked as in a scenario where a grade was awarded perfectly then the rest is up to other aspects of the student which is an ideal situation.

In a paper by Tomkinson et al. (2011), they stated that a strategy taken to prevent the incorrect marking of student assessments specifically in the undergraduate sector is to have multiple markers and then use the average as the actual result for the student. Tomkinson et al. (2011) also talks of a "halo" and "horns" effect where the halo effect refers to where "supervisors give higher marks than the written work merits because they have been aware of the effort and thought processes" and the horns effect is when "a dilatory student produces a dissertation of greater merit than the supervisor has been led to expect." Also, another problem pointed out by Tomkinson et al. (2011) with using second or third markers is that they may not have sufficient knowledge concerning the subject area and as such cannot mark the student to the degree of accuracy expected. These are the kinds of problems which would be solved by the introduction of my system as there won't be a need for a second or third evaluator.
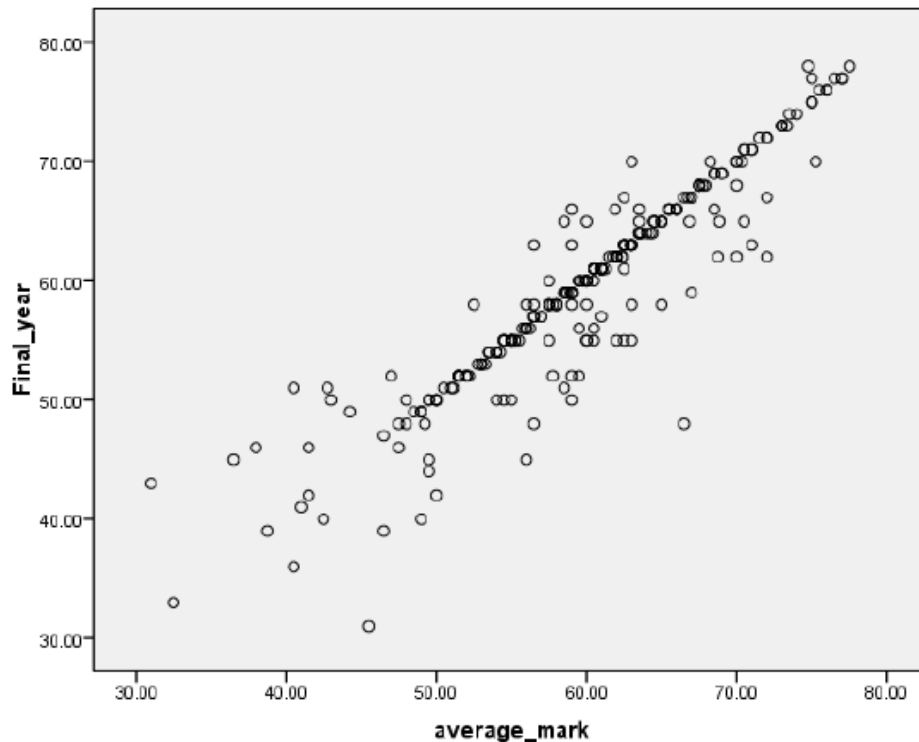

Figure 1. Final marks against average of first and second markers

The results of Tomkinson et al. (2011) research showed that the second markers who are generally not the lecturers of the course award lower marks on average while the first markers award higher but when the average has been collected it shows consistency with each of their markings.

AI which is short for Artificial Intelligence and it mainly deals with creating systems that can imitate intelligent human functions. In particular NLP is one of the most challenging aspects of AI due to the fourfold nature of it; Speech Recognition, Syntactic analysis, Information Extraction and Discourse Analysis. It mainly deals with human language which can have different meanings stemming from the same sentence due to differences in things like punctuation and two completely different sentences can have the same meaning. Artificial Intelligence can be incorporated into the learning process and also make it more efficient and better. Luckin (2017) stated in her article that "AI is a powerful tool to open up the 'black box' of learning." AI has already been employed in the assessment of essays known as AES (Automated Essay Scoring) with the most popular of this category being IntelliMetric. These systems have high reliability and are able

to grade and provide feedback within seconds something teachers cannot do. This category of software is unique in the sense that it wants to understand the meaning of the text and as such determine if it relates to the question asked and grade it appropriately.

The system I propose is one that has a set of questions and a sample answer to the questions set by the educator and the system will perform a Compares between them to determine their degree of similarity. This removes the need for the system to fully understand the text and the hidden contexts but it does take depth away from the system. The proposed system will also have the added bonus of evaluating if the educator actually has an understanding of the assignment or assessment set as they will have to provide their own answers to the question. This way it will remove inherent issues like the one Dikili (2006) stated in her paper regarding PEG (Project Essay Grader); "Since PEG used indirect measures of writing skill, it was possible to trick the system, i.e., writing longer essays."

Although this would be an optimal solution to solve some of the multiple issues with traditional grading it does come with its own drawbacks. One of the issues is what (Luckin, 2017) highlighted in her article; she states AI can be very costly to implement especially when looking at massive systems that can handle the grading of multiple students across multiple subjects. Also, to complete a system that could have such capabilities it would need to be backed by some national or large corporation as well as have access to the core details of the student and the curriculum as well as being able to determine which areas the student is struggling with for future development. Luckin (2017) stated "this suggests an annual budget of US$600 million per year for a complex AI project. It therefore seems reasonable to suggest that a country, such as England, might need to spend the equivalent of US$600 million (£500 million) per year to make AI assessment a reality for a set of core subjects and skills."

Previously the only work done on using systems for assessments was in the area of multiple-choice questions and similar shown by researches by Ana et al. (2013) and Boussakuk et al. (2021). But Wilson et al. (2000) proposed the creation of the system BEAR. This stands for Berkeley Evaluation and Assessment Research the proposed system would be capable of understanding the curriculum of the school and in turn accessing the students throughout the year and through their work. They used the IEY (Issues, Evidence and You) developed by SEPUP (Science Education for Public Understanding Project) Course to test their program. The program had a grading scheme which was adapted form SOLO Taxonomy by Biggs et al. (1982). According to Wilson et al. (2000), their grading scheme is a system ranging from 1 which is "an answer with only one correct aspect to it" to 4 which is a perfect answer by the student.

In Wilson et al. (2000) paper they also express the need conform to standards of fairness which they state as including "Consistency and Unbiasedness." This also correlates with some of the main problems with traditional grading that I outlined in the earlier parts of the literature review. Ultimately this study was just for a system to evaluate the grading by teachers and where the students need help as opposed to a standalone AI that can grade the students by itself and also provide feedback on where the student got it wrong.

Evidence and Tradeoffs (ET) Variable

Score	*Using Evidence:* Response uses objective reason(s) based on relevant evidence to support choice.	*Using Evidence to Make Tradeoffs:* Response recognizes multiple perspectives of issue and explains each perspective using objective reasons, supported by evidence, in order to make choice.
4	Response accomplishes Level 3 AND goes beyond in some significant way, such as questioning or justifying the source, validity, and/or quantity of evidence.	Response accomplishes Level 3 AND goes beyond in some significant way, such as suggesting additional evidence beyond the activity that would further influence choices in specific ways, OR questioning the source, validity, and/or quantity of evidence & explaining how it influences choice.
3	Response provides major objective reasons AND supports each with relevant & accurate evidence.	Response discusses at least two perspectives of issue AND provides objective reasons, supported by relevant & accurate evidence, for each perspective.
2	Response provides some objective reasons AND some supporting evidence, BUT at least one reason is missing and/or part of the evidence is incomplete.	Response states at least one perspective of issue AND provides some objective reasons using some relevant evidence BUT reasons are incomplete and/or part of the evidence is missing; OR only one complete & accurate perspective has been provided.
1	Response provides only subjective reasons (opinions) for choice and/or uses inaccurate or irrelevant evidence from the activity.	Response states at least one perspective of issue BUT only provides subjective reasons and/or uses inaccurate or irrelevant evidence.
0	No response; illegible response; response offers no reasons AND no evidence to support choice made.	No response; illegible response; response lacks reasons AND offers no evidence to support decision made.
X	Student had no opportunity to respond.	

Figure 2. Wilson et al. (2000) scoring guide

_____

3. Methodology

3.1 *Conceptual research framework*

The research will be split into three main phases and are as follows. Figure 3 gives a diagrammatic representation of the phases.
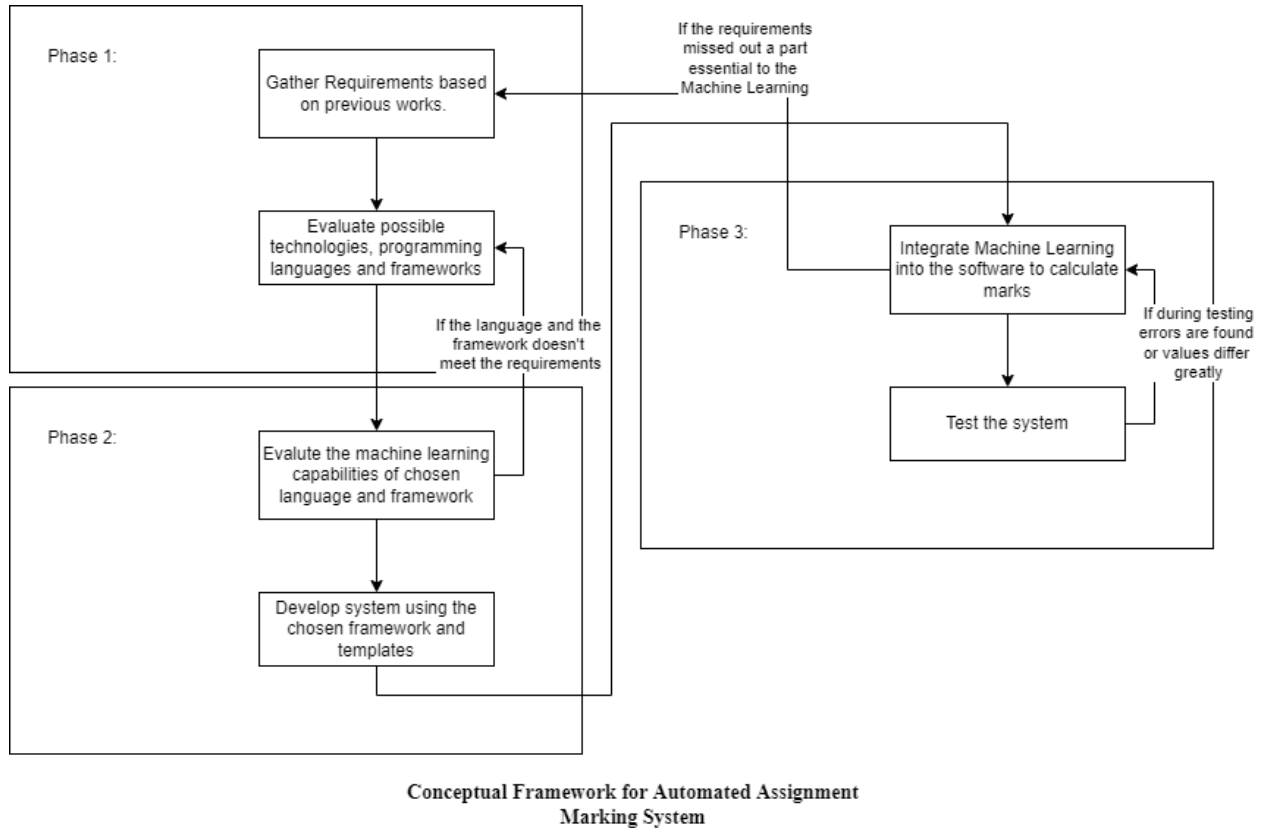

Conceptual Framework for Automated Assignment Marking System

Figure 3. Conceptual Framework.

3.1.1 *Phase 1*

Phase 1 is the Requirement gathering phase and the framework choosing phase. During this phase I gathered the basic requirements of an assignment submission system without the evaluation of marks and then I added the Machine Learning to calculate marks automatically. The next aspect was to choose the programming language (Python, C++, C# or any other) to use and the platform the application would be deployed to as well as the specific technology to use to create the application (I had to pick between any of the following Winforms, React with .NET, WPF, Xamarin, Django, Android Studio with JAVA or with Kotlin).

3.1.2 *Phase 2*

Phase 2 is the evaluation of the chosen language and technology and whether or not it has the Machine Learning Libraries needed for the completion of the software. Owing to this I chose not to use any other language but Python due to the vast number of Libraries available and the versatility of it and the framework I chose was Django so I can create a web application which could be widely available. During this phase I also developed the system to be deployed using some templates and developing original parts to use to fulfil the requirements gathered in phase 1.

_____

3.1.3 *Phase 3*

Phase 3 is the part where I added the Machine Learning code to the already complete system to compute the results. For this phase I used the Spacy ML Library which is mainly a library for Natural Language Processing and has already trained models and also is dependent on other Models such as NumPy and Doc2Vec and others to compute its result. After this aspect was done, I moved into testing of the system and any errors encountered were fixed the errors which stemmed directly from the requirement lacking a part were solved by modifying the requirements and then going through the phases again until the end phase has been reached.

3.2 *Application methodology: Waterfall model*

The methodology used for the creation of the application was the Waterfall model. This is a particular implementation of the software development life cycle that focuses on sequential development like a waterfall. This methodology has each phase completely wrapping up before the next phase begins. The waterfall model is highly dependent on a lot of work being done at each stage as there is no going back. It is efficient for small projects and can provide an effective release date.

3.3. *Stages in the waterfall model*

The waterfall model has 5 stages or phases and team members usually work independently on each stage though phases have to be completed in sequential order.

The stages namely Requirements, Design, Implementation, Verification and Testing are discussed below with the aid of Figure 4.

Figure 4. The waterfall model

3.3.1 *Requirements*

The project requirements have to be gathered and understood before any work can be commenced. The project requirements will be obtained from the stakeholders. This will be presented in the form of a document which contains details about each stage and also other important bits such as timelines, cost, risks and the success rates.

_____

3.3.2 *Design*

The developers are required to design a technical solution based on the requirements. This is where things like scenarios, layouts and data models. This also where the scope of the project is identified.

3.3.3 *Implementation*

After the design has been completed this is where the technical implementation begins using hardware and software technologies. This is where the coding is done based on the requirements and specifications. Changes are usually minimal in this stage but if big changes need to be implemented then it's to go back to the design phase.

3.3.4 *Verification or testing*

Testing is done before the product or service can be released to the public, testing techniques such as white hat testing, black box testing and the like are done at this stage.

3.3.5 *Deployment and maintenance*

This is when the software is actually out for use and this is also when plans for future versions are made.

3.3.6 *Advantages and disadvantages of the Waterfall Model*

The waterfall modes are advantageous for many reasons one is because it helps system designers to find errors during the design and analysis stage and saves them the trouble in the implementation and testing stage. Another is that the cost and time for the software to be delivered can be estimated. Progress can be followed because the end of each stage is a milestone to reach. New developers can understand the project easily due to the extensive requirements document. Also, since it is not as iterative it is completed faster since the stakeholders aren't adding new unnecessary features.

The disadvantages though are that if the project is big, it will take a longer time to complete than the agile methodology or the V model. Clients don't usually know everything they want from a software at the start so they prefer to ask for changes during development and new features later down the road. This methodology also means the clients are not involved in the design and implementation stages. Finally, the biggest issue with this methodology is if one phase is delayed all the other phases are delayed.

3.3.7 *Why do I use the Waterfall Model*

I selected the waterfall model for the sole purpose that it is used for small programs that do not require large amounts of requirements. I also use it as this is a project that offers a clear intention of how the project will be done and how the software should look like from the get go. Also, the requirements of the program won't change as I go ahead and create the program. All the stages of the program would be outlined from inception to implementation. If this program does go on to be used by other institutions it would be better to use another model such as the V model.

_____

3.4 *Functional requirements*

These are the things the system should do and the features the software should provide in order to gather these requirements. The stakeholders in this system are the users, the students and the lecturers.

Below are the functional requirements that have been obtained from the stakeholders mentioned above:

1. Allow students to login and logout.
2. Allow students to sign up.
3. Allow lecturers to login and logout.
4. Allow lecturers to sign up.
5. Allow lecturers to upload assignments.
6. Allow lecturers to upload sample answers.
7. Allow lecturers to view students' submissions.
8. Allow students to submit assignments for lecturers' course.
9. Allow lecturers to view grade of the assigned work.
10. Allow lecturers to delete assignments and submissions.
11. Allow lecturers to create and delete course.
12. Allow lecturers and students to edit details.
13. Allow users to email for newsletters and updates.

3.5 *Non-functional requirements*

These are the quality aspects of the system. They are not as essential to the functionality of the system but they are preferred.

1. Security

The system should be secure so whoever uses the system knows their details are not compromised. The system should be private and it should have integrity. It should be able to stop hackers from obtaining information about a customer from a system and then using it to exploit said customer for malicious or financial gain.

2. Usability

The system should be friendly and allow the users to interact with it comfortably. If a product has a lot of features and is not usable then the users will choose to go with another artifact or service that is easier to use and can be remembered intuitively.

3. Availability

The system should be available when the users need to use it and there shouldn't be downtime with any of the servers or the database. It should also be available on any of the web platforms.

4. Performance

The system should be responsive and fast and provide quick results. The system should not slow down unnecessarily or keep the users waiting and the system should be able to handle a high number of requests without downtime.

5. Accuracy

The systems should be able to provide accurate results for the user and the lecturer to see. The results should be consistent with what a teacher would award a student in every scenario and should provide consistency

6. Fault tolerance

The systems should be able to handle errors within the system and send the appropriate message to the users. The errors should be minimized from the developers and care should be taken so the system doesn't crash outright under unforeseen circumstances.

7. Efficiency

The system should be as efficient as possible and consume the least resources possible. The system should be able to use the least threads possible on the server to ensure that it runs optimally.
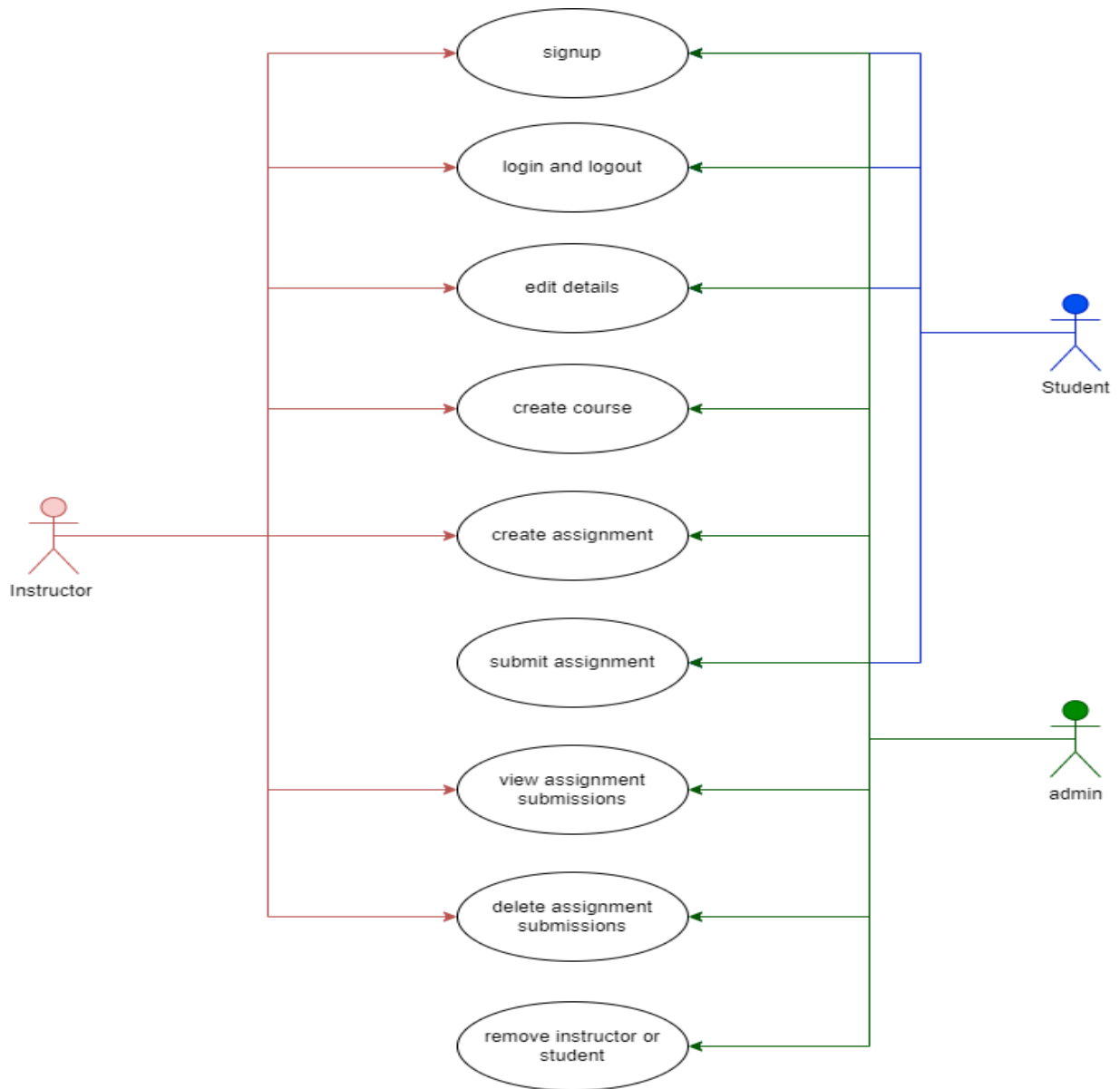
8. Cost

The system is a free to use systems so it is very cost effective from that standpoint. But generally, software's shouldn't be too expensive for the end user but should be appropriately priced so the users would be happy and the developers appropriately paid.

3.6 *Tools used*

1) Visual Studio Code, https://code.visualstudio.com : A lightweight Integrated Development Engine developed by Microsoft for use with all languages and frameworks. It runs on windows, Mac OS and Linux as well.

2) Postman, https://www.postman.com: This is an API platform built for developers to create and test API's before they are used in development.

3) Browsers: Chrome and Edge to test the final outlook of the page.

4) Python, https://www.python.org: This is the main language that was used for the development of the application.

5) GitHub, https://github.com/: This is a code repository, where you can find templates and other projects by other developers as well as examples. It can also be used for collaboration between developers and project management.

6) Django, https://www.djangoproject.com/: This is the main framework used to develop this application, Django is based on python and supports dynamic pages as well as SQLite; this is a database written in languages and can be embedded in frameworks. It is present in android studio as well.

7) Stack Overflow, https://stackoverflow.com/: This is an open platform for developers to post their issues and get answers as well as pointers on what to do to solve their problems.

8) Quora, https://www.quora.com/: This is a question-and-answer social platforms designed for all sorts of questions.

9) Windows, https://www.microsoft.com/en-gb/windows/?r=1: This an Operating System, the most popular operating system in the world. It developed by Microsoft and it has multilingual support and was one of the first OS to come with a GUI (Graphical User Interface).

10) spaCy, https://spacy.io/: This is a very fast and easy to use software library for advanced natural language processing founded 7 years ago. It also be used in other frameworks such as TensorFlow and pyTorch.

_____

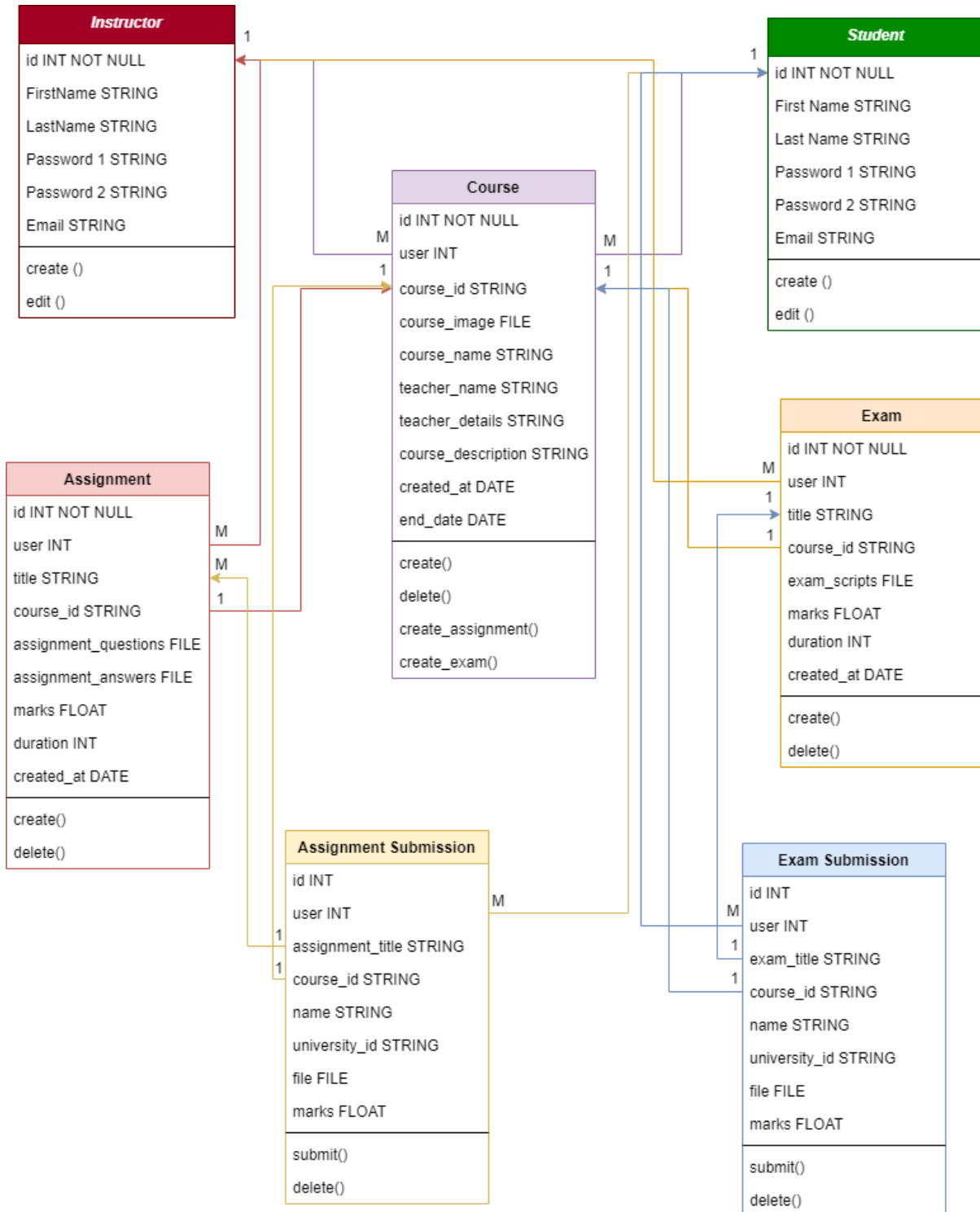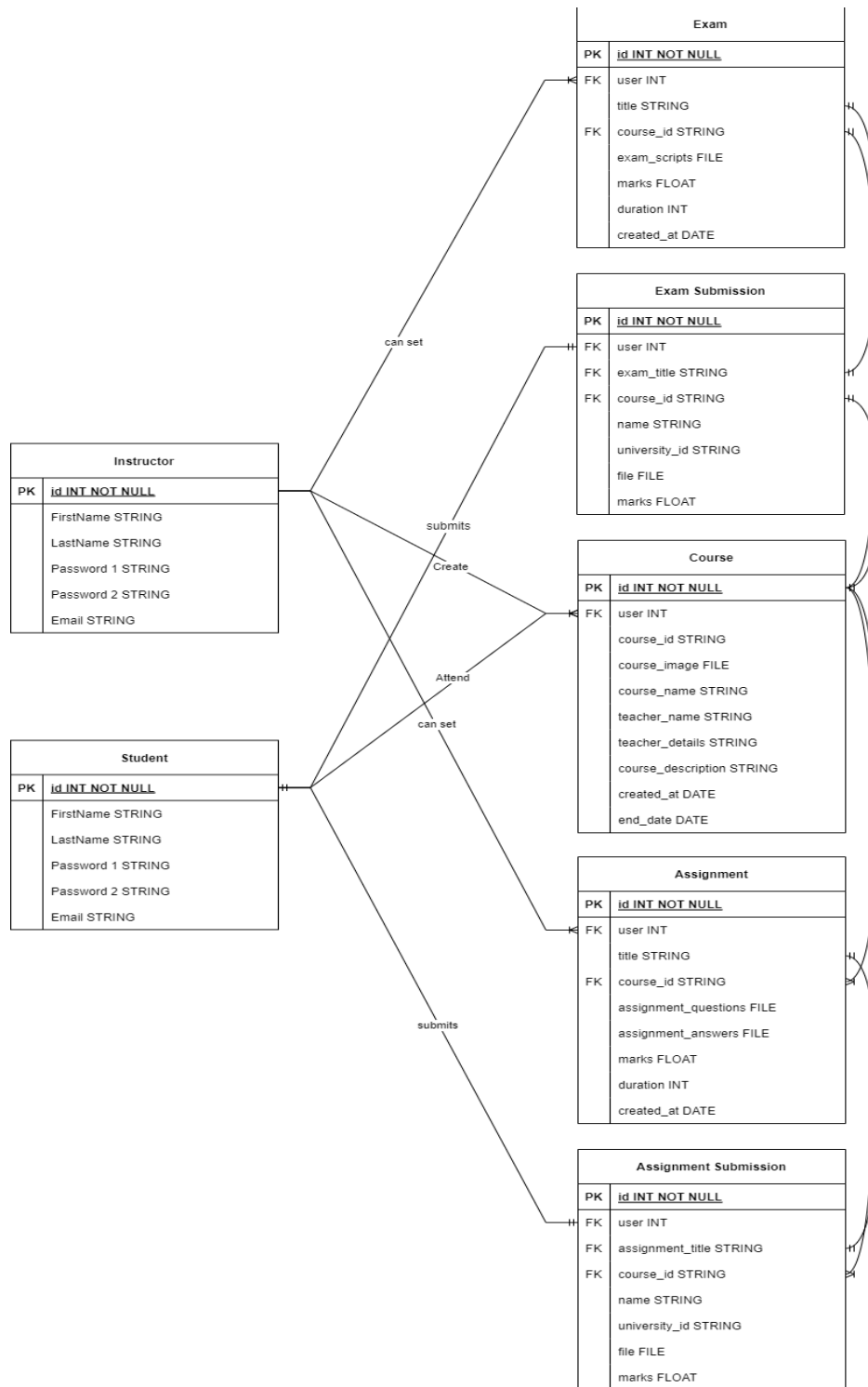### 3.7. *Diagrams*

3.7.1 *Use case diagram*

3.7.2 *Activity diagram*
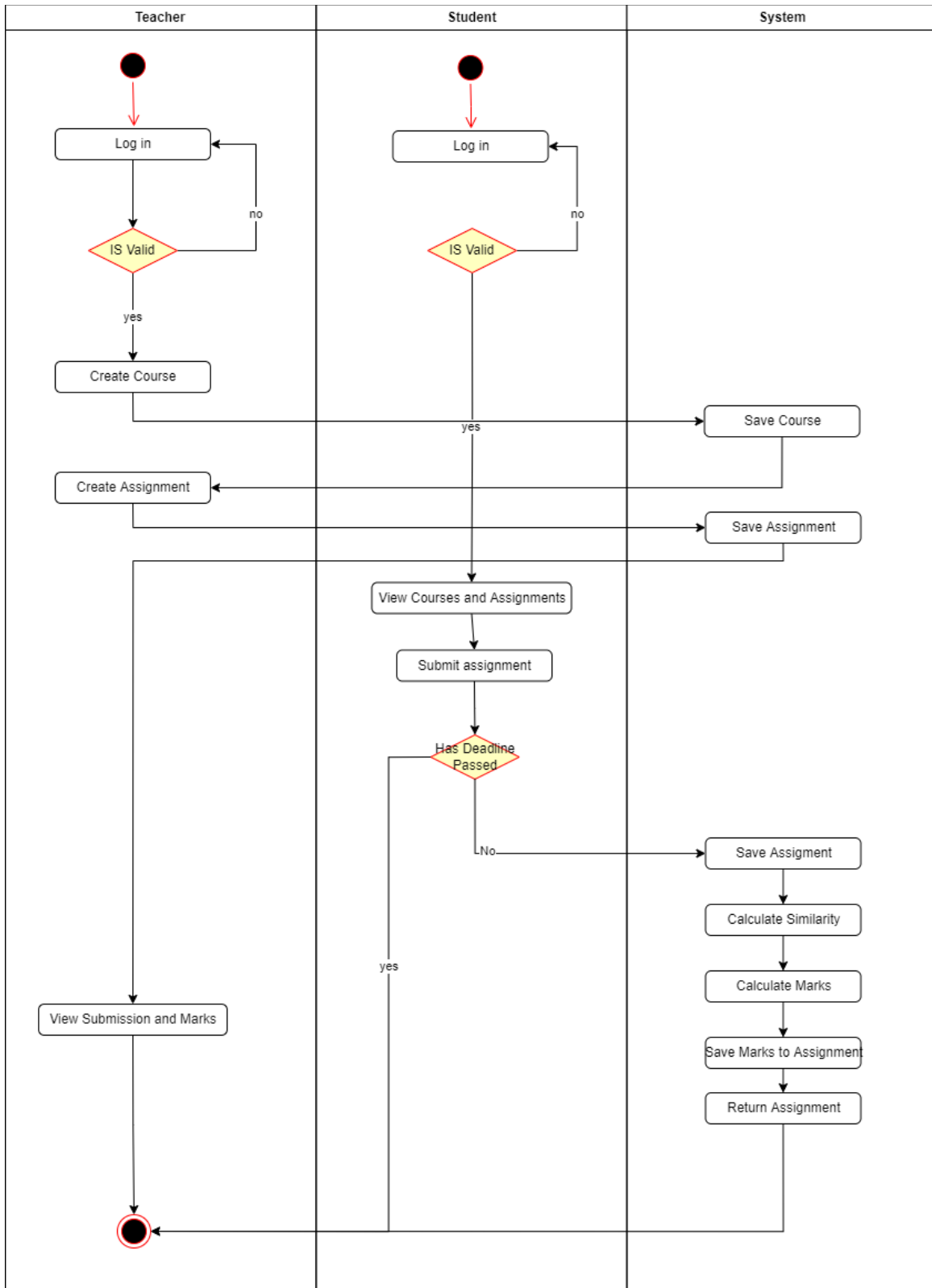
3.7.3 *Class diagram*

4. Implementation

4.1 *Entity relation diagram*

Exam

PK	id INT NOT NULL
FK	user INT
	title STRING
FK	course_id STRING
	exam_scripts FILE
	marks FLOAT
	duration INT
	created_at DATE

Exam Submission

PK	id INT NOT NULL
FK	user INT
FK	exam_title STRING
FK	course_id STRING
	name STRING
	university_id STRING
	file FILE
	marks FLOAT

Instructor

PK	id INT NOT NULL
	FirstName STRING
	LastName STRING
	Password 1 STRING
	Password 2 STRING
	Email STRING

Course

PK	id INT NOT NULL
FK	user INT
	course_id STRING
	course_image FILE
	course_name STRING
	teacher_name STRING
	teacher_details STRING
	course_description STRING
	created_at DATE
	end_date DATE

Assignment

PK	id INT NOT NULL
FK	user INT
	title STRING
FK	course_id STRING
	assignment_questions FILE
	assignment_answers FILE
	marks FLOAT
	duration INT
	created_at DATE

Student

PK	id INT NOT NULL
	FirstName STRING
	LastName STRING
	Password 1 STRING
	Password 2 STRING
	Email STRING

Assignment Submission

PK	id INT NOT NULL
FK	user INT
FK	assignment_title STRING
FK	course_id STRING
	name STRING
	university_id STRING
	file FILE
	marks FLOAT

can set
submits
Create
Attend
can set
submits

_____

4.2 *Testing*

Testing was conducted with answers from Quora on the question: What is design thinking? The link to the page is placed below and the results of the program are specified below.

https://www.quora.com/What-is-Design-Thinking

The results of the program as well as the answers are put in Table 1 below.

Table 1.

	Answers	Score Awarded
Sample Answer	Design Thinking is following a human-centric approach while dealing with any problem. First identifying the problem which is affecting humans and then finding the solution. The very common issue with human behavior is that we always tend to directly jump on the solutions. We don't make efforts to find the real cause behind a problem (Context). Design Thinking helps us in finding the real cause keeping in mind three main aspects: People, Technology, and Business.	50
1.	There are various problems in this world. Many of them are complex and some of them are simple. We need to solve those problems. Since many solutions are available to a single problem but out of those many, we need to figure out the best one. The best solution is generally termed as the innovative and that is out of the box. Thinking divergently instead of conversantly and to come up with a creative idea that might not exist for a problem that works well for the real problem, is all about Design thinking. In one line we can state "Design Thinking is a process for creative problem solving". Design thinking is a human-centered approach to innovation that draws from the designer's toolkit to integrate the needs of people, the possibilities of technology, and the requirements for business success."	28.786405880065214
2.	There are diverse issues on this world. Many of them are complicated and a number of them are simple. We want to remedy the ones issues. Since many answers are to be had to an unmarried hassle however out of these many, we want to determine out the pleasant one. The pleasant answer is commonly termed because the progressive and this is out of the box. Thinking divergently rather than conversantly and to provide you with an innovative concept that may not exist for a hassle that works properly for the actual hassle, is all approximately Design wondering. In one line we are able to state "Design Thinking is a method for innovative hassle solving". Design wondering is a human-focused technique to innovation that attracts from the designer's toolkit to combine the desires of people, the opportunities of technology, and the necessities for enterprise success."	23.029742123351877

3.	There are many problems in the world. Some of them are complex and some of them are simple. We need to find a way to solve those problems. Since many solutions are available for one problem but among those many, we need to find out the best one. The best solution is usually innovative and out of the box. Creative thinking is all about coming up with ideas that are different from the ones that are already out there. By thinking divergently, you can create new solutions to problems that work better than the ones that are currently being used. Design Thinking is a process for solving creative challenges. It is a human-centered approach that uses the design tools of a designer to integrate the needs of people, the possibilities of technology, and the requirements of business success.	28.73406602683103
4.	There are numerous issues on the planet. Some of them are mind boggling and some of them are straightforward. We really want to figure out how to take care of those issues. Since numerous arrangements are accessible for one issue yet among those many, we want to figure out the best one. The best arrangement is normally creative and out of the container. Innovative reasoning is tied in with concocting thoughts that are not the same as the ones that are now out there. By thinking differently, you can make new answers for issues that work better compared to the ones that are presently being utilized. Configuration Thinking is an interaction for settling imaginative difficulties. A human-focused approach utilizes the plan instruments of a creator to coordinate the necessities of individuals, the conceivable outcomes of innovation, and the prerequisites of business achievement.	26.526140094197046
5.	Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged.	4.3467948056326335

5. *Results and findings*

5.1 *Summary and conclusion*

After the development of the program and testing we see that the data in numbers 1 through to 4 entered even though they are small paragraphs more or less mean the same thing as the sample answer but the spaCy model returned the scores to range from 23-28 out of 50 marks. This means the answers were half correct according to the models. These same answers which when compared by a human are more or less the same thing. The only exception here is the 5th answer with 'Lorem Ipsum' this was thrown to intentionally test the system and the system responded appropriately and awarded it 4.34 out of 50 marks. This means the above answers that got more than half marks are considered somewhat similar according to the system. This correlates with Luckin (2017) who stated estimated that "an annual budget of US$600 million per year" would be needed for a complex AI project. The spaCy model is free and open Source so there could be issues whether the NLP Libraries are really as advanced as they should be. In its defense though the similarity function was created to handle single sentences instead of large paragraphs.

_____

5.2 *Recommendations*

My first recommendation to improve the validity of this software is to use the largest model offered by spaCy, due to space constraints I used the medium model which was 400MB for the libraries alone and a further 1.4GB for everything else added on.

The next recommendation is to use a dedicated paid API designed to handle paragraphs so the development of the system can be quick.

My third and final recommendation is that proper research over a period of time should be put into creating an AI model specifically for marking essays and one that can do it by itself without the aid of anything else such as sample answers or marking scripts.

5.3 *Limitations*

This research will be limited by the fact that the system can only handle the marking of short sentences accurately and not long paragraphs. The system is also limited by the fact that it can only mark with the aid of the marking scheme and not without it so it is not a truly intelligent model in that regard. The system will be limited by the fact that it can only compare answers in English and not in other languages or disciplines. Another limitation is the fact that the documents have to only contain the answers and no other additions such as a cover page for the system to work effectively.

Acknowledgements

References

Adams, W. L. (1932). Why teachers say they fail pupils. *Educational Administration and Supervision*, XVIII, 594-600.

Ana, P., & Tawo Bukie, P. (2013). *Design and implementation of online examination administration system for universities*. https://www.researchgate.net/publication/327075106.

Archer, A., & B. McCarthy (1988). *Personal biases in student assessment*. https://doi.org/10.1080/0013188880300208

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.

Bloxham, S. (2009). *Marking and moderation in the UK: false assumptions and wasted resources*. https://doi.org/10.1080/02602930801955978

Boussakuk, M., Bouchboua, A., El Ghazi, M., & El Bekkali, M. (2021). *Designing and developing e-assessment delivery system under IMS QTI ver.2.2 specification*.

Dikli, S. (2006). Automated essay scoring. *Turkish Online Journal of Distance Education-TOJDE, 7*(1), Article: 5.

Elliot, S. (2000a). *A study of expert scoring and IntelliMetric scoring accuracy for dimensional scoring of grade 11 student writing responses* (RB-397). Newtown, PA: Vantage Learning.

Elliot, S. (2000b). *A true score study of IntelliMetric accuracy for holistic and dimensional scoring of college entry-level writing program* (RB-407). Newtown, PA: Vantage Learning.

Elliot, S. (2001a). *About IntelliMetric* (PB-540). Newtown, PA: Vantage Learning.

Elliot, S. (2001c). *Applying IntelliMetric Technology to the scoring of 3rd and 8th grade standardized writing assessments* (RB-524). Newtown, PA: Vantage Learning. 61.

Elliot, S. (2002). *A study of expert scoring, standard human scoring and IntelliMetric scoring accuracy for statewide eighth grade writing responses* (RB-726). Newtown, PA: Vantage Learning.

Elliot, S. (2003a). *A true score study of 11th grade student writing responses using IntelliMetric Version 9.0* (RB-786). Newtown, PA: Vantage Learning.

Elliot, S. (2003b). *Assessing the accuracy of IntelliMetric for scoring a district- wide writing assessment* (RB-806). Newtown, PA: Vantage Learning.

Elliot, S. (2003c). *How does IntelliMetric score essay responses?* (RB-929). Newtown, PA: Vantage Learning.

Elliot, S. (2003d). IntelliMetric: From here to validity. In M. D. Shermis & J. C. Burstein (Eds.). *Automated essay scoring: A cross disciplinary approach*. Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, M. R. Jr. (2022). *Student assessment in teaching and learning.* https://cft.vanderbilt.edu/student-assessment-in-teaching-and-learning/.

Hill, P., & Barber, M. (2014). *Preparing for a renaissance in assessment*. London: Pearson Education.

James, D. A., & Gabunada Lambating, J. (2001). *Validity and Reliability in Assessment and Grading: Perspectives of preservice and in-service teachers and teacher education professors*. Conference of the American Educational Research Association (Seattle, WA, April 10-14, 2001).

Kuisma, R. (1999). *Criteria referenced marking of written assignments*. https://doi.org/10.1080/0260293990240103

Kukich, K. (2006). Beyond automated essay scoring. *The Turkish Online Journal of Distance Education*. https://www.semanticscholar.org/paper/Beyond-Automated-Essay-Scoring-Kukich/66a03e431c858ed3dd00d25773e3a2c9b5528e6c.

Luckin, R. (2017). *Towards artificial intelligence-based assessment systems*. http://dx.doi.org/10.1038/s41562-016-0028

Luckin, R., & Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. Pearson Education, London.

McKinstry, B., Cameron, H., Elton, R., & Riley, S. (2004). Leniency and halo effects in marking undergraduate short research projects. *BMC Medical Education*, *4*, 28-28.

Nottingham, M. (1988). *Grading practices – Watching out for land mines.* NASSP.

Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE—Life Sciences Education*, *13*, 159-166.

Tomkinson, B., & Freeman, J. (2011). *Problems of assessment*. Conference: *International Conference on Engineering Education* (ICEE-2011). Belfast, Northern Ireland. https://www.researchgate.net/publication/273455702_Problems_of_Assessment.

Willey, K., & Gardner, A. (2010). *Improving the standard and consistency of multi-tutor grading in large classes*. https://www.uts.edu.au/sites/default/files/Willey.pdf.

Wilson, M., & Sloane, K. (2010). From principles to practice: An embedded assessment system. *Applied Measurement in Education*, *13*, 118-201. https://doi.org/10.1207/S15324818AME1302_4

Appendix

Home Page

View Course Page

View Assignment

Create Assignment Page

Edit Profile Page

View Assignments Page

Delete Assignments Page

Submit Assignments Page

Signup Page

Django Admin

Django administration

Site administration

AUTHENTICATION AND AUTHORIZATION		
Groups	+ Add	✏ Change

CORE		
Assignment submissions	+ Add	✏ Change
Assignments	+ Add	✏ Change
Courses	+ Add	✏ Change
Exam submissions	+ Add	✏ Change
Exams	+ Add	✏ Change

Recent actions

My actions

✗ VUG/CSC/19/1000
Assignment submission

✗ VUG/CSC/19/2000
Assignment submission

✗ VUG/CSC/19/3000
Assignment submission

✗ VUG/CSC/19/4000
Assignment submission

✗ VUG/CSC/19/5000
Assignment submission

✗ VUG/CSC/19/1000
Assignment submission

✗ VUG/CSC/19/4000
Assignment submission

✗ VUG/CSC/19/3000
Assignment submission

✗ VUG/CSC/19/2000
Assignment submission

✗ VUG/CSC/19/5000
Assignment submission