**1. Introduction**

1.1 Introduction

A graph of a student

Description automatically generatedIn Sri Lanka students in their final collegiate year (age 18-19) sits for General Certificate of Education (Advanced Level). Each year over 240,000 students qualify to sit for the G.C.E (Advanced Level) exams. This exam assists as an entrance requirement for many Sri Lankan state universities and many other universities. The students are expected to study two years in school before participating for the exams. The exams are conducted in three languages Sinhala, English and Tamil and diverse over four fields (Math, Science, Commerce, Art, and Technology). Each student must choose three subjects under one field. According to the University Grant Commission Sri Lanka, in 2017, only 19.25% of qualified candidates were admitted to undergraduate courses, The percentage slightly decrease to 18.98% in 2018.In 2019, it increased to 22.98%. Its further increase reaching 22.58% in 2020 and 25.39% in 2021.

Figure 1- (Sri Lanka University Statistics 2022, 2022)

The above statistics showcase the intense competition for this examination among students each year.

Among all the General Certificate of Education (Advanced Level) subjects, the project discussed in this report emphases exclusively on the subject chemistry. Students of both bio science and physical science sits for the chemistry paper. Annually, approximately 75,000 students sit for the G.C.E (Advanced Level) Chemistry exam. In 2021, 76,851 sits for the exam and shows 29.58% failed percentage, in 2022, 75,612 students sat for the exam and 28.80% failed the exam. The exam for the subject chemistry involves two papers. Paper one and paper two are held on two days, paper one prior to paper two. The first paper consists of 50 MCQs (Multiple Choice Questions) that contributes 50 marks directly to the final grading of the subject. The second paper consists of 4 structured questions and 4 essay questions that carries total 1000 marks which will be later taken out of 50 marks. The 50 MCQs on the paper 1 can be categorized in to five categories as inorganic, organic, environmental & industrial, physical, and general chemistry. As mentioned before the first paper is held prior to paper two and carries 50 marks directly to the final grading. Hence the students are keen on scoring more on paper one to secure their results early and more easily.

Students that sit for the exam, quizzes themselves using MCQ past paper questions and model paper questions with improving paper 1 results in mind. Often students attend to private tutors for this but to achieve better results self-studying is unavoidable. Thought self-studying plays a crucial role in improving one’s result, without proper guidance this may not be effective.

1.2 Problem Definition

As previously mentioned, the paper 1 of chemistry exam consists of questions from five subject areas (Organic chemistry, Inorganic chemistry, Physical chemistry, Environmental & Industrial and General chemistry). Many students engage in answering more mcq questions to recognize the subject areas they lack knowledge in and excel in. By identifying these areas they’ll be able to grasp what subject matters that need more of their attention. Though tutors provide some assistance in recognizing these subject areas, the onus of improving the knowledge on subject matters clearly lies on the students themselves. The process of recognizing subject areas that need more attention tend to be time consuming but pivotal. Recognizing the areas, they lack knowledge affect directly to the performance enhancement process. This helps students to fill their knowledge gaps and improve understanding. Likewise, if the students are well prepared for the exam, the stress and anxiety they feel will be reduced facilitating optimal performance in the exam. Often students go through around 5-6 amount papers to recognize these subject areas. After recognizing, students tend to practice themselves with questions from those lacking areas. For this they either manually filter those subject questions from past papers or search for books that has subject vise categorized questions in them. Both methods can be time consuming. Time is precious for these students as they have two more subject that need the same amount attention and self-studying.

In conclusion, its crystal clear that to address these challenges an online platform that carter students need is required. A platform that offers a diverse number of questions in each subject category, can recognize the students’ weak areas efficiently and connect teachers and students when the students need further guidance. Hence, the platform discussed here intentions to improve the overall learning experience of students and academic performances.

A diagram of a computer process

Description automatically generatedChemX is a system that’s designed to facilitate students in improving their academic performance in the subject of chemistry. The fundamental functions of the system are as below. The system includes three users: students, teachers, and administrators. It was previously mentioned that the subject chemistry can be divided into five categories but here in order to reduce the complexity of the project only three categories are considered. As users the students can take part in question answering once they logged into the system. The answers will be passed to the ML model where the correct number of questions in each category is identified. Afterward he results are passed back hence displayed through a chart to the student. The teachers are able to add new questions to the database. Once the question is submitted, it will be monitored by the administrator and added to the database. The administrator can manage users as well as manage the question database.

A diagram of a login and user manual

Description automatically generated

Approve new MCQ



1.3 Project Objectives

1. Building up confidence in students when solving exam questions.
2. Recognizing the areas where students excel and struggle in when answering questions.
3. Creating a question bank with questions from all the categories.
4. Help students to improve the effectiveness of their self-studying time.
5. Help students to manage time in their studies.
6. Provide students with instant feedback.
7. Help students to access different model questions from different tutors.
8. Help students to reach and get feedback from peers and teachers anytime they want.
9. To assist students in tracking their progress with each round of questions.
10. Developing a user-friendly/ easy to use platform for both students and teachers.
11. Helps teachers to reach out to students more often and check on their progress.
12. Facilitates students with continuously growing question bank.
13. Encouraging collaboration between students and teachers through the chat feature.
14. Helps teachers to improve their teaching skills.
15. Create an online community where chemistry questions can be discussed.
16. This project can be used as a blueprint in creating MCQ applications for other subjects.

**2. System Analysis**

2.1 Facts Gathering Techniques

Two types of data were gathered during the phase of fact gathering. Primary data was gathered using interviews and questioners and secondary data was gathered using existing systems and research papers. A screenshot of a computer survey

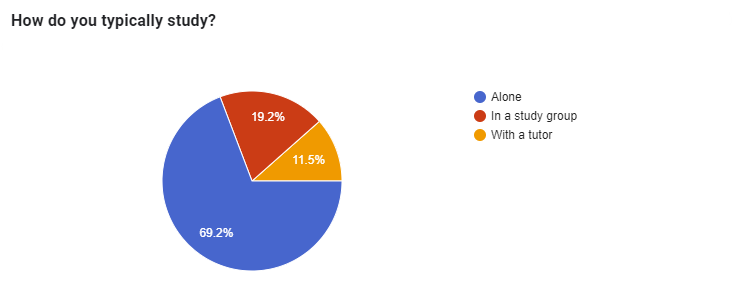
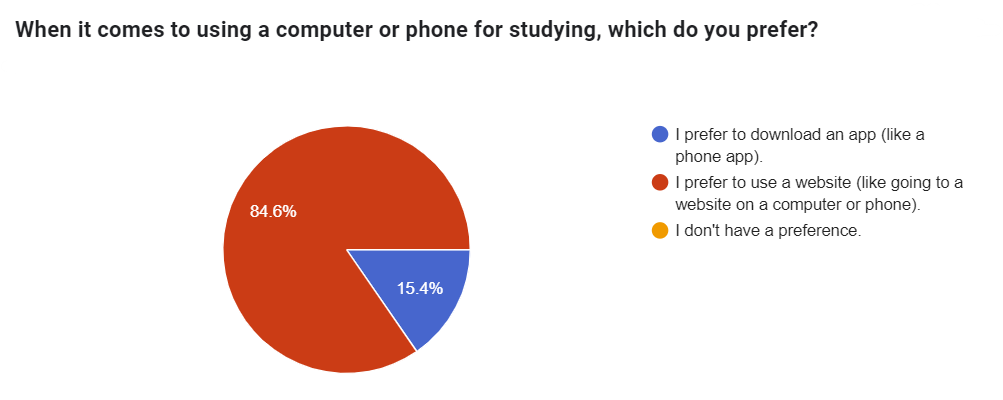
Description automatically generatedA screenshot of a computer

Description automatically generatedThere are two groups of users for this project namely students and teachers. Primary data was gathered for both type of users separately. Accordingly, online questioners were used to gather data from school students.

A screenshot of a computer

Description automatically generated

A pie chart with text on it

Description automatically generatedShown above are the questionnaire and its analysis. Through the online questioner it was proven that in general most of the students tend to study alone. This data concludes that it’s crucial for the project focuses on helping students that study alone. Majority expressed that they haven’t studied with the aid of a quiz application before and expressed that it would be helpful to track one’s progress over mcq rounds. When asked what’s the preferred tool used to help studying the majority choose the web applications while the others choose using mobile applications. This data on the preferred studying tool helps to develop the project in students preferred application type.

A white paper with black text

Description automatically generatedThis highlights the use of technology with school students for their studies. Also stresses the need for an adoptive, user-friendly educational platform for students that provides an effective learning environment. In conclusion the analysis underscores the growing use of technology in the education sector, which is a good sign for the development of this project. Concurrently, Interviews were conducted to gather facts from teachers.

2.2 Existing System

In addressing the use of technology in the field of educations the use of web-based application attest to be a transformative solution. Akhil Bharadwaz et al., 2023 in his research discusses about My Exams, a web-based application for examinations. This system is made to assist both teachers and students. It is mentioned that the students are registered to the system by the institute and the students are expected to get their login details through the institute. The teachers can conduct online tests, assessments, and quizzes through this application. Also, the teacher can create update delete exam questions while the administrations have the ability to manage students. The answers of the students who goes through the exam process are evaluated and the stored in the database. This is later used by teachers and administrators for student evaluate. The system ensures accurate results and lessen time constrains by the automated evaluation process.

Navik et al (2021) in their work talks about a web-based system for mcq examinations. This is designed for institutes and colleges. The main goal is to replace the traditional exam methos of manually answering exam questions with an automated system. The students must register before accessing the system. The system consists of three types of users: students, instructors, and administrator. The instructions can manage exams. The administrator has the ability to add, remove and update users of the system. The main aim of the system of this system is to is to make the exam process simpler and reduce human error in marking. The paper concludes stating that the system provides a user-friendly, secure online platform for conducting exams.

Aaboud, 2023 in his thesis mentions about a Quiz game application for medical students in Finland. Aaboud, 2023 emphasizes that using the surge in digital technology he hopes to build an innovative solution for medical students goes through extensive amount of studying and preparations. Hence the thesis explores the ability to offer an engaging study environment to medical students in hopes of making their study time efficient and effective. The system includes diverse of medical categories, facilitating exam questions and tests for each category. The system is designed to be simple and user friendly in order to aid students in concentrating on subject matters rather than technical details. Once the students logged into the system, they’re presented with two options “Game” and “Study materials”. The first option offers the users to access the quiz rounds while the latter allow the user to go through the study materials. The system also includes an admin interface which enables admins to constantly add new categories, questions and answers to the system keeping it up to date.

Nuha Alruwais and Zakariah (2023) in their paper explains about a system that uses Machine learning techniques for evaluating assessments of students. The paper discusses the usage of various prominent ML algorithms for this. The dataset contains past assessment data of students and two of these datasets were used. The system predicts both fail/pass and grade of a student. For these seven classifiers were used (SVM, MLP, LR, RF, DT, GBM, and GNB). The highest accuracy was given by the Gradient Boosting Machine algorithm (GBM) with the dataset 1, an accuracy of 98%. The paper also mentions that there was no significant change when datsset-2 is used. Hence concluded the paper stating how efficient the use of GBM in evaluating students’ performance.

Usage of multi class classification model in predicting grades of students is discussed by Bujang et al., 2021.The paper discussed how the semester end grade of a student is calculated through the Machine learning model. The dataset used here contain 1282 instances of student first semester grades. Out of all the first semester courses only two courses are selected for the research. For data preprocessing the students are categorized into five groups according to their grade. (Exceptional, Excellent, Distinction, Pass, and Fail). It is stated that hence the data distribution show an imbalance Synthetic Minority Oversampling Technique (SMOTE) is used. Wrapper and filter-based methods are used for feature selection. The paper uses five widely known algorithms: Random Forest, Logistic Regression-Nearest Neighbor, Naïve Bayes, Support Vector Machines. The selected models are trained and tested with the datasets. The accuracy of each model is evaluated. The results for the classification are presented using data visualization techniques to helps understand the overall performance of each model. Bujang et al., 2021 concludes the paper by presenting the findings and conclusions of this research.

The research done by Engr. Sana Bhutto et al., 2020 talks about using supervised learning to predict students’ academic performance. The paper emphasizes how the use of algorithms plays a crucial role in enhancing the educational sector. The papers mention that by use of machine learning its possible to gain a better understanding of students’ performances, learning patterns of students, foresee outcomes of students and grouping students according to their learning patterns. Furthermore, its stated that by identifying these patterns and relationships learning institutes will be able to minimize the annual dropout rate of students. The data set for this research includes 500 records obtain from e-leaning system. Null data is removed in data cleaning reducing the dataset to 480 records. The goal of the model is to predict the future performance of students for this the marks of each student are grouped three as good, average, and bad. For classification two prominent algorithms Logistic Regression and Support Vector Machine are used. The paper concludes the research stating that the algorithm support vector machines showed an accuracy of 78% outperforming logistic regression (71% accuracy).

Luo, Chen and Han, 2020 conduct the research on predicting students’ final grade on blended courses using their online behavior data. Luo, Chen and Han were able to analyze the online behavioral patterns of students and categorize them into five types. These patterns in behavior are later used to predict their final grade outcome of the blended course. The data for this research includes online behavioral data and final grades of students totaling up to 229795 records. As mentioned before through clustering the records were categorized into five categories as Inactive, Low-active, Assignment, Video, and High-active. The Random Forest Algorithm was used to build the prediction model. The model gave good results for four blended courses out of five courses. In conclusion through the research, it was revealed that the students in blended courses have different online patterns.

Understanding the integration of the aforementioned papers contributed largely to building the discussed project, ChemX. Inspired by the finding gathered from the above papers the project has undergone noteworthy enhancements. Both Nuha Alruwais and Zakariah (2023) and Bujang et al., 2021 practice around five algorithms to get the best model for their system. Considering the positive outcomes mentioned in the two ChemX employed the same technique and practice five algorithms later choosing the one with the best accuracy. Furthermore, in his paper Aaboud, 2023 mentions how he enables admin to constantly update the question database. This technique too was used in the Mcq system-ChemX enabling teachers to constantly add questions to the database.

2.3 Use case diagram

A screenshot of a computer

Description automatically generated

2.4 Drawbacks of the existing system

**3. Requirements Specification**

3.1 Functional Requirements

1. Registration and login for the users.

-The admin, teachers and students should be able to register and then login to the system

1. Quiz for the students.

-The students should be able to take quizzes with 15 mcq in each round from three chemistry categories.

1. Evaluation on each round.

-It should be possible for the students to check their progress in charts after finishing each round

1. Next Question recommendation.

-The system should recommend the next question category to the students at the end of a question round depending on the performance.

1. Tracking the progress.

-Students should be able to track the progress of all mcq rounds

1. Chatroom for students and teachers

-Users should be able to discuss questions with each other in the chatroom.

1. Submit new questions.

-Teachers should be able to submit new questions to the database.

1. Check progress of students.

-Teachers can check progress of relevant students

1. Manage questions.

-Admin manages the questions submitted by teachers and add appropriate questions to the database.

1. Manage users.

-Admin manages the users by removing inappropriate users.

3.2 Non-Functional Requirements

1. The details of the users especially the passwords should be stored securely.
2. System should prevent unauthorized accesses.
3. The system should load pages without a significant delay.
4. The database should be able to accommodate the growing number of questions and users.
5. System should be able to respond to the user actions within 3 seconds.
6. The interfaces should be responsive.
7. The user interfaces should be user-friendly.
8. The feedback from the ML model should be given faster.
9. The ml model should give correct outcomes.
10. The system should be easy for users to use.

3.3 Hardware / Software Requirements

1. Front end

-HTML, CSS, JavaScript

1. Server

-Node.js, Express.js, MongoDB atlas

1. Machine Learning

- Jupyter Notebook, python

1. Other

-Git, Figma, Visual studio code

**4. Feasibility Study**

4.1 Operational Feasibility

4.2 Technical Feasibility

4.3 Outline Budget

**5. System Architecture**

5.1 Class Diagram of Proposed System

5.2 ER Diagram

5.3 High-level Architectural Diagram

5.4 Networking Diagram (Optional)

**6. Development Tools and Technologies**

6.1 Development Methodology

A diagram of a computer process

Description automatically generated

6.2 Programming Languages and Tools

6.3 Third Party Components and Libraries

6.4 Algorithms

The dataset for the Machine Learning model consists of 100 records gathered from ten students. Each row represents a MCQ round. Each round includes 15 MCQs, total correct numbers of MCQ, Correct number of mcq in each category, Category with highest number of correct answers and category with lowest number of correct answers [figure 6.4.1]. Since this data set has big amount pf variable/columns and lesser number of records/rows this data set is considered as wide data set.

A table with numbers and letters

Description automatically generated with medium confidence

[figure 6.4.1].

A screenshot of a computer

Description automatically generatedData cleaning was carried out by checking null values in the dataset. Two rows were detected with null values, and they were dropped from the dataset reducing the row number from 100 to 98. Next the columns *‘Student\_ID’* and *‘Round’* were dropped from the dataset as the columns doesn’t have any effect on the targeted labels. The cleaned data was saved as a separate file [figure 6.4.2].

[figure 6.4.2]

Next the dataset is split as features and labels. The features consist of answers for questions from 1-15 and total number of correct answers while the labels consist of Total correct answers on each category and Category with highest number of correct answers and category with lowest number of correct answers [figure 6.4.3]

A close-up of a white background

Description automatically generated

[figure 6.4.3]

A screenshot of a computer code

Description automatically generatedFollowing the dataset is split into 60% Training set ,20% test dataset and 20% validation set.

[figure 6.4.3]

**7.Discussion [Max of 1 Page]**

Overview of the Interim Report

Summary of the Report

Challenges Faced

Future Plans / Upcoming Work

**8.References**

1. Administrator (2022). Sri Lanka University Statistics 2022. [online] Ugc.ac.lk. Available at: [https://www.ugc.ac.lk/downloads/statistics/stat\_2022/Chapter%202.pdf](https://www.ugc.ac.lk/downloads/statistics/stat_2022/Chapter%202.pdf%20) [Accessed 23 Feb. 2024].
2. Anon, (2022). Performance of Candidates G.C.E.(A/L) Examination – 2022 (2023). [online] Available at: <https://doenets.lk/images/resources/STAT/final%20POC-AL-2022-web_1694447239752.pdf.>
3. Anon, (2021). Performance of Candidates G.C.E.(A/L) Examination – 2021. [online] Available at: <https://doenets.lk/images/resources/STAT/AL2021POC-Rev-1_1671814563246.pdf.>
4. Akhil Bharadwaz, B., Sasidhar, B., Vali, Bhagawati, R. and Jain, S. (2023). My Exam: A Web-Based Application for MCQ Type Online Examination. IRE Journals |, [online] 6, pp.2456–8880. Available at: <https://www.irejournals.com/formatedpaper/1704476.pdf.>
5. Navik, J., Sawant, A., Saini, S. and Shinde, V. (2021). PERFORMANCE ANALYSIS SYSTEM USING WEB BASED SOLUTION FOR MCQ TYPE ONLINE EXAMINATION. Certified ISSN, [online] 6. Doi:<https://www.oaijse.com/VolumeArticles/FullTextPDF/758_9.PERFORMANCE_ANALYSIS_SYSTEM_USING_WEB_BASED_SOLUTION_FOR_MCQ_TYPE_ONLINE_EXAMINATION.pdf>
6. Aaboud, M. (2023). Developing a quiz application for medical students. [online] theseus.fi. Available at: [https://www.theseus.fi/bitstream/handle/10024/795396/Aaboud\_Maroua%20Hibatollah.pdf?sequence=2](https://www.theseus.fi/bitstream/handle/10024/795396/Aaboud_Maroua%20Hibatollah.pdf?sequence=2%20) [Accessed 8 Mar. 2024].
7. Nuha Alruwais and Zakariah, M. (2023). Evaluating Student Knowledge Assessment

Using Machine Learning Techniques. Sustainability, [online] 15(7), pp.6229–6229. doi:

<https://doi.org/10.3390/su15076229.>

1. Bujang, A., Selamat, A., Ibrahim, R., Ondrej Krejcar, Herrera-Viedma, E., Fujita, H. and Nor (2021). Multiclass Prediction Model for Student Grade Prediction Using Machine Learning. IEEE Access, [online] 9, pp.95608–95621. doi:<https://doi.org/10.1109/access.2021.3093563.>
2. Engr. Sana Bhutto, Isma Farah Siddiqui, Qasim Ali Arain and Anwar, M. (2020). Predicting Students’ Academic Performance Through Supervised Machine Learning. [online] doi:<https://doi.org/10.1109/icisct49550.2020.9080033.>
3. Luo, Y., Chen, N. and Han, X. (2020). Students’ Online Behavior Patterns Impact on Final Grades Prediction in Blended Courses. [online] doi:<https://doi.org/10.1109/eitt50754.2020.00034.>
4. ‌
5. ‌
6. ‌
7. ‌
8. ‌
9. ‌
10. ‌

‌