

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

# **Learnverse: A Computational Performance Prediction Tool in a Learning Environment**

Group 13 Project Proposal Document by

Inuka Rodrigo - 20210508  
Lisara Gajaweera - 20211029  
Sanjula Maneth - 20211481  
Haneek Ahamad - 20200665

Supervised by  
Mr. Prasan Yapa

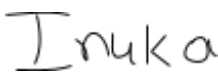



Submitted in partial fulfillment of the requirements for the BSc (Hons) in Artificial Intelligence And Data Science degree at Robert Gordon University.

**October 2022**

© The copyright for this project and all its associated products resides with  
Informatics Institute of Technology

## Declaration

We hereby certify that this project proposal and all the artifacts associated with it is our own work, and it has not been submitted before nor is currently being submitted for any degree program.

Student Name	Student ID	Signature
Inuka Rodrigo	20210508	
Lisara Gajaweera	20211029	
Haneek Ahamed	20200665	
Maneth Subawickrama	20211481	

I have read the project proposal, and it is in accordance with the approved university project proposal outline.

Signature of the supervisor:

Date:

# Table Of Contents

---

<b>Declaration</b>	<b>2</b>
<b>Introduction</b>	<b>5</b>
<b>Problem Domain</b>	<b>5</b>
<b>Problem Statement</b>	<b>5</b>
<b>Research Motivation</b>	<b>6</b>
<b>Existing work</b>	<b>6</b>
<b>Research Gap</b>	<b>7</b>
<b>Contribution to the Body of Knowledge</b>	<b>8</b>
<b>Technological Contribution</b>	<b>8</b>
<b>Contribution to domain</b>	<b>8</b>
<b>Research Challenge</b>	<b>9</b>
<b>Research Questions</b>	<b>9</b>
<b>Research Aim</b>	<b>10</b>
<b>Research Objective</b>	<b>10</b>
<b>Project Scope</b>	<b>11</b>
<b>In-scope</b>	<b>11</b>
<b>Out-scope</b>	<b>11</b>
<b>Feature Prototype</b>	<b>12</b>
<b>Resource Requirements</b>	<b>14</b>
<b>Hardware Requirements</b>	<b>14</b>
<b>Software Requirements</b>	<b>14</b>
<b>Data Requirements</b>	<b>14</b>
<b>Skill Requirements</b>	<b>14</b>
<b>Methodologies</b>	<b>15</b>
<b>Research Methodology</b>	<b>15</b>
<b>Development Methodology</b>	<b>15</b>
<b>Project Management Methodology</b>	<b>16</b>
<b>Risk associated with project and mitigation plan</b>	<b>16</b>
<b>Gantt Chart</b>	<b>17</b>
<b>Deliverables</b>	<b>18</b>
<b>Reference</b>	<b>19</b>

## List Of Tables

<b>Table 1 : Existing Work .....</b>	<b>6</b>
<b>Table 2 : Research Objectives.....</b>	<b>10</b>
<b>Table 3 : In Scope.....</b>	<b>11</b>
<b>Table 4 : Out Scope.....</b>	<b>11</b>
<b>Table 5 : Research Methodologies.....</b>	<b>15</b>
<b>Table 6 : Risk and Mitigation Plan.....</b>	<b>16</b>
<b>Table 7 : Project Deliverables.....</b>	<b>18</b>

## List of Figures

<b>Figure 1 : Feature Prototype (Process 1).....</b>	<b>12</b>
<b>Figure 2 : Feature Prototype (Process 2).....</b>	<b>13</b>
<b>Figure 3 : Gantt Chart.....</b>	<b>17</b>

## Introduction

Learnverse is a multimodal educational system which has the capacity of improving a student's performance by generating model papers based on how he/she performs in previous papers. Our system mainly targets students who are doing advanced level ICT in Sri Lanka. Our system generates model papers with 50 multiple-choice questions that assess all of the skills required for each lesson in the GCE A/L ICT syllabus in Sri Lanka.

## Problem Domain

Around 316,737 ([3.5 Number Literate and Literacy Rates by Census years & Sex, 2015](#)) students take a/l exam each year in Sri Lanka. Among them, most students cover the a/l syllabus from schools and private institutes. Most private institutes conduct classes under two categories. Mass classes which consist of minimum 250 students and group classes which consist of maximum 50 pupils. Very minimal number of students go for group classes and individual classes so that teachers can give their full focus on students while most students don't get attention from their teachers. Therefore, students struggle hard to figure out which skills they need to improve and which lessons they must additionally focus on.

Even if a student realizes which lessons or skills they are weak at, it would be a time-consuming task for them to find out questions from past papers and model papers related to those specific lessons. When the a/l exam gets closer, some students may need to communicate with the teacher very often to clarify their doubts about lessons. But it's not possible to meet the teacher regularly.

## Problem Statement

This study focuses on improving the performance of GCE AL students who study ICT as a main subject. The students studying for examination have to study a vast range of subject areas related to the ICT subject. With the limited time available, making the study time productive is a must to score well in this competitive examination. With the busy schedule of studies and extracurricular activities, students are confronted with difficulties with time management and study planning. When the exam is getting closer, students need to focus on areas where they can score more and the areas where they are weak at.

Without concentrating on the study areas and skills that need more focus, scoring in the examination is very challenging. Therefore, making this process less stressful and more straightforward should be done. Using a model to acquire questions which have a higher chance of appearing in the examination, helps students to score more in the a/l exam. From

this method, students who tend to score less and fail the examination can easily improve themselves and score higher in the examination.

This approach is usually taken by the tuition class teachers. But the accuracy of their questions are questionable. And most of the time those questions are merely based on the examination, not on the areas that should be focused more by the specific student individually. Therefore, building a model for this task is a must to make better predictions and to increase students' performance.

## Research Motivation

Each year more than thousands of students face the Advanced Level examination in Sri Lanka. Here only a small percentage of the students achieve higher results while others reach for lower average or failing grades. This research project focuses on guiding the students for better results by analyzing the performance of the students and recording the progress of each student. Unlike in physical classes ,any student who is lacking behind shall be recognised by the system itself and will assist those students to work more on the lessons which they are struggling with. Up until now there has not been a system which has been able to do these tasks, therefore it is safe to assume that this project shall benefit both students and teachers breaking the boundaries of traditional classrooms.

## Existing work

The G.C.E Advanced Level examination of srilanka is one of the most competitive examinations in the world where a lot of hardwork and dedication is involved. Yet there is no specific automated method to assist the traditional learning system. However there are some few research projects that have been brought up which perform similar tasks in the following research project.

Citation	Technology / Algorithm	Advantages	Limitations
<a href="#">(Vachev et al., 2022)</a>	Matching the question with the answer using the MASK token.	Automated MCQ paper generation to minimize time wastage.	The subject contents were taken as the basic data input for the preparation of the question papers.
<a href="#">(Xu and Liao, 2021)</a>	MVC Framework	Generating Random and fair Question papers for the students.	Low Question quality.

<a href="#">(Rahim et al., 2017)</a>	Genetic Algorithms	Preparing high quality exam questions.	Questions are not specified according to the individual students' performance.
<a href="#">(Wu et al., 2020)</a>	Genetic Algorithms and Deep Knowledge Tracing	Able to predict students' performance based on the past questions he/she has answered.	Students are not able to get questions based on their preference.

## Research Gap

Students who are facing A/L exams most often identify their weak areas through teachers. The skills that students do not have vary from one student to another. Therefore teachers must give attention to each and every student individually to figure out which skills they are lacking, and which lessons they must focus on. It is not possible for a teacher to pay attention to each pupil. A system is therefore needed to identify students' weak points and improve them.

Many research studies have investigated the implementation of generating automated MCQ papers. However, few studies have examined the generation of MCQ papers on the basis of students' weak competencies. Therefore, we propose a system which generates MCQ papers based on students' past performance.

Another thing is that existing systems do not produce MCQ papers to evaluate specific competencies.

For example, a student wants to do more questions from the lesson "logic gates" even though he's good at that lesson. But existing systems aren't capable of identifying students' requirements and producing papers which only consist of questions related to "logic gates".

In this system, students' requirements will be collected via the information provided through chatbot and generate model papers based on those requirements.

## **Contribution to the Body of Knowledge**

### **Technological Contribution**

Our educational system will use supervised machine learning algorithms such as multivariate regression to predict lessons where the student is weak and use genetic algorithms to generate questions along with a correct answer and additional incorrect answers (distractors). An Angular web application will be created using HTML, CSS , JS to demonstrate the algorithm. Flask REST API will be used to interact with datasets stored in databases.

Furthermore, the analysis of the performance of the students shall be processed with python programming with the help of python libraries.

### **Contribution to domain**

With the use of above-mentioned technologies, we will design a user-friendly system which is capable of making model papers based on skills students are weak at.

For example, a student will do an ICT model paper and once he completes it, the system identifies that the student has answered questions related to ” operating systems” subject incorrectly. So next time, he will get a paper that includes more questions related to "operating system" lesson.

Planning to use an AI chatbot to gather information from students regarding lessons they want to study further. This will help students to develop their skills and knowledge on related areas.

For example, a student can communicate with an AI chatbot and let it know that he wants papers related to logic gates. Therefore, the system itself generates MCQ based model papers which only consists of questions related to “logic gates”.

Apart from that, this project also provides teachers to track the progress of their students by answering model papers given by the system. Unlike any other system, here each students’ progress is analyzed and displayed to the teacher so that he/she can see which lessons have to be revised in the future.



## Research Challenge

- Due to the unavailability of direct research papers for the project, the researchers have to make their own datasets from referring to the G.C.E Advanced Level Examination past papers. Further assistance can be taken to overcome this challenge by referring to the past papers categorized by [Pesuru Publications Ltd.](#)
- Implementing a user-friendly interface for the students to assure the acceptance by clients. In order to overcome this, the researchers have decided to use a web interface, since the interface can be styled into a more preferable scope using CSS styling and other web developing software programs.
- Establishing a synchronized online classroom for the teacher and his/her students for setting the paper class is a major challenge that the researchers have to face. In order to deal with this, many research papers have to be gone through to identify the specific technologies. The technologies used here might lead to a publication since the research papers are taken by different journals such as Computers & Education([www.sciencedirect.com](http://www.sciencedirect.com), n.d.).
- One of the major challenges would be in preparing the customized model papers for the students by means of their performance. Here, the concepts of machine learning shall be used to train data into the system.
- Implementing the chatbot to understand the requirements of the student and assist them when in need would be a troublesome task considering the time availability of the due time for the whole project.

## Research Questions

RO1: Which algorithm is best for predicting students' results based on their performance?

RO2: How does an AI bot understand a student's requirements and provide a solution for them?

RO3: Which methodology can be used for building an online platform for students and teachers?

RO4: How can students' performance be used to generate MCQ based papers for them?

## Research Aim

This proposed system will identify students' weak competencies based on how he/she performs in model papers and generate new model papers including more questions related to areas that student is weak at. Furthermore, AI chatbot will be used to get information about lessons which students require for further improvement and generate special papers only consisting of questions related to those specific lessons.

## Research Objective

Research Objectives	Explanation	Learning Outcome
Problem Identification	<p><b>RO1:</b> Identifying the best algorithm to model the question paper according to the students' performance.</p> <p><b>RO2:</b> Selecting the best computational strategy for improving the accuracy of the students' exam result prediction.</p> <p><b>RO3:</b> Designing a lenient web interface for the users to work.</p> <p><b>RO4:</b> Building the dataset, which consists of questions categorized according to the lesson.</p> <p><b>RO5:</b> Implementing a communication platform for the teacher and students, and analyzing the performance of the student</p> <p><b>RO6:</b> Selecting a suitable technology to implement an AI chatbot for the students.</p>	LO1
Data Gathering and Analysis	<ul style="list-style-type: none"> <li>Self set data gathered from G.C.E A/L I.C.T past papers.</li> <li>Data gathered from previous Research papers from Google scholar, IEEE, ResearchGate etc.</li> <li>Journals.</li> </ul>	LO1, LO3
Research Design	Quasi- Experimental Design will be used as the Research design as it aims to establish a cause and effect relationship between dependent and independent variables ( <a href="#">Thomas, 2020</a> ).	LO4, LO3
Implementation	<p>Different Implementations in this research project are as follows:</p> <ul style="list-style-type: none"> <li>Implementation of a web application to answer modeled MCQ papers for the students based on their performance.</li> <li>Student Profile to check their progress and predicted grades for the A/L exam.</li> </ul>	LO2, LO3, LO4

	<ul style="list-style-type: none"> <li>AI chatbot for assisting students with academic work.</li> <li>Implementation of virtual classrooms to evaluate the class performance.</li> </ul>	
Testing and Evaluation	Surveys and questionnaires shall be used for testing and evaluation of the research project. Surveys are used to get the opinions and feedback from the students and teachers in order to evaluate the quality of the research project ( <a href="#">QuestionPro, 2019</a> ).	LO2, LO4

## Project Scope

### In-scope

No	Description
1	Predict the model paper formats from past papers.
2	Provide a set of practice questions related to the requested study area of the subject.
3	Create a virtual classroom where the teacher can view the performance of students.
4	Predict a student's grade based on the model paper marks.

### Out-scope

No	Description
1	Creating the model for many subjects and languages.
2	Generate entirely new model questions
3	Creating the virtual classroom which can manage all the other subjects as well.
4	Predicting the Z score of the student in the examination.

## Feature Prototype

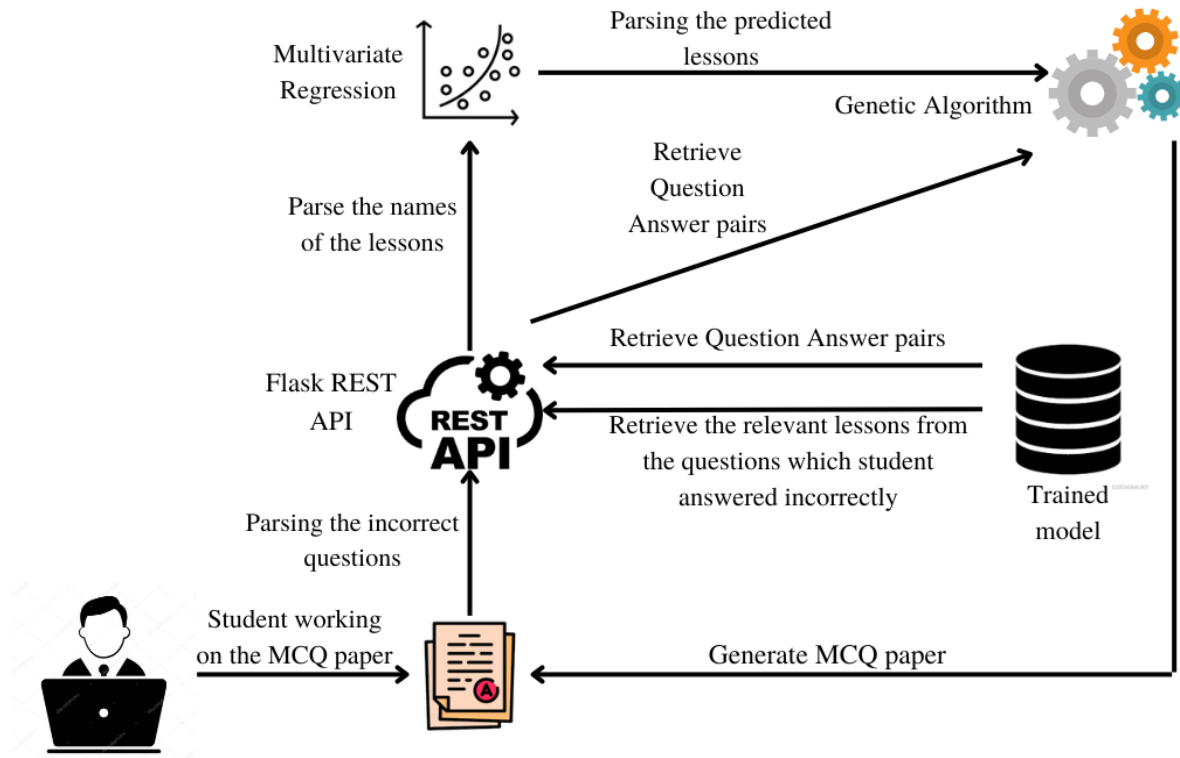


Figure 1: Feature Prototype (Process 01)

### Process 01

- Pre - requisite: 1st paper is generated manually
1. Student will work on the MCQ paper through the web site created.
  2. After that questions which he answered incorrectly will be parsed to the Flask REST API.
  3. Retrieve the names of the lessons for the questions which the student answered incorrectly from the trained model.
  4. Parsing the names of the lessons to the multivariate regression model.

Multivariate Regression model - This model has the records of students' past performance (Relative lessons of the questions to which student obtained incorrect answers in previous papers).

5. Multivariate Regression model will predict the lessons students' weak at and parse the names of the lessons to the genetic algorithm.
6. Genetic algorithm will retrieve question - answer pairs along with other incorrect answers from the trained model via Flask REST API.
7. Generating MCQ papers from the questions retrieved from the trained model. These questions were added based on the weightage given to each lesson.

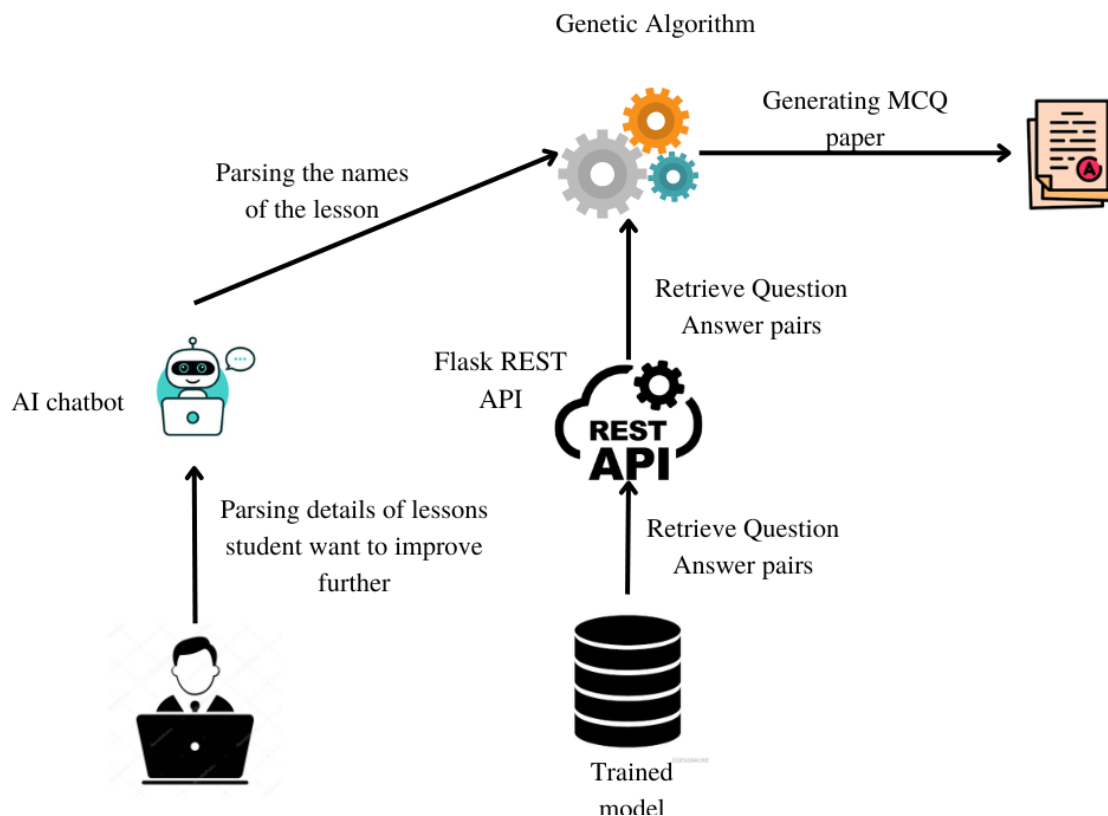


Figure 2: Feature Prototype (Process 02)

## Process 02

1. Student parses the names of the lessons which he needed for further improvement through an AI chatbot.
2. AI chatbot will gather the information and parse it to the genetic algorithm.
3. Genetic algorithm will generate the MCQ paper based on the lessons student asked using the questions retrieved from the trained model.

## Resource Requirements

### Hardware Requirements

- **CPU (Intel Core i7 7th generation processor or high)**- To get more powerful and high performance.
- **16GB RAM or high** - To train heavy algorithms.
- **Storage (minimum 256GB SSD / 1TB HDD )** - To store a large amount of data.

### Software Requirements

- **Python** - The primary language used to create the proposed system is Python, which is particularly user-friendly in terms of error handling and library support.
- **PyCharm Enterprise** - used for developing proprietary and commercial software.
- **SQLite**– To manage databases and servers.
- **HTML**- To structure the webpage.
- **CSS** – used to style and layout web pages.
- **JavaScript** – used for scripting the webpage.
- **MS Word** – To write documents.
- **PHP** – to develop dynamic and interactive websites.
- **TensorFlow** - For pre-processing and training the model.
- **Windows Operating System** – To handle huge computational functionalities .
- **SPSS** - To run linear regression with multiple dependant variables (Multivariate regression)

### Data Requirements

Since we are focusing on the advanced level ICT exam in Sri Lanka, there are no suitable datasets found for our project. Therefore, our team must create our own dataset that is suitable for a/l ICT students.

### Skill Requirements

- Searching for information
- Time management
- Problem solving
- Report writing
- Critical thinking
- Planning and scheduling

## Methodologies

### Research Methodology

Research Philosophy	<p>The researchers have selected positivism as the research philosophy because the data used for the process of the project are factual.</p> <p>Here the work of the researchers is limited to data collection and interpretation which means that they are independent in the following Research Project.</p>
Research Approach	The research follows an inductive approach. Starting with research questions and identifying the patterns of the students' performance in a test, the system theorizes the students' under-achieving lessons and generates untested model papers.
Research Strategy	Quantitative Questionnaires is used as a research strategy, because the analyzing of the students' performance can be done only after the student answers a considerable amount of question papers.
Research Choice	This Research will be using a mono method where only one method shall be used in gathering data.
Time zone	Cross-sectional Time zone is used because the research goes through a specific data set only a single time.

### Development Methodology

This project uses Agile development as it can be used to decompose the entire project into sprints based on user stories (stories are software features in the project) using scrum, which is also an Agile framework. Agile model is capable of testing and delivering software features sprint wise and makes the project more manageable. Iterative development methodology can be used to deliver the project to users. Requirements of this project can be dynamic, therefore Agile is the most suitable software methodology as it provides a better way of managing the changing requirements. Object-oriented analysis and design(OOAD) is used to design our system, as it uses iterative methodology to refine and extend our design. And since the requirements of our project are dynamic, this will be the most suitable technical approach.

## Project Management Methodology

To do the project management, Scrum methodology is used, which is a derivative of Agile framework. A Gantt chart is used to get the big picture of the overall project. Click up and Asana web based applications are used to integrate group work. GitHub and Google Collab applications will be used to store, track and collaborate on the project as a group.

## Risk associated with project and mitigation plan

Risk Item	Severity	Frequency	Mitigation Plan
Hardware Unavailability	5	4	Use cloud based/virtual computation
Loss of project data	5	1	Use cloud based storage and backup frequently
Accuracy of the predictions	3	3	Run tests and improve the system
Requirement changes	4	2	Finalize the aims of the project before implementing system



## Gantt Chart

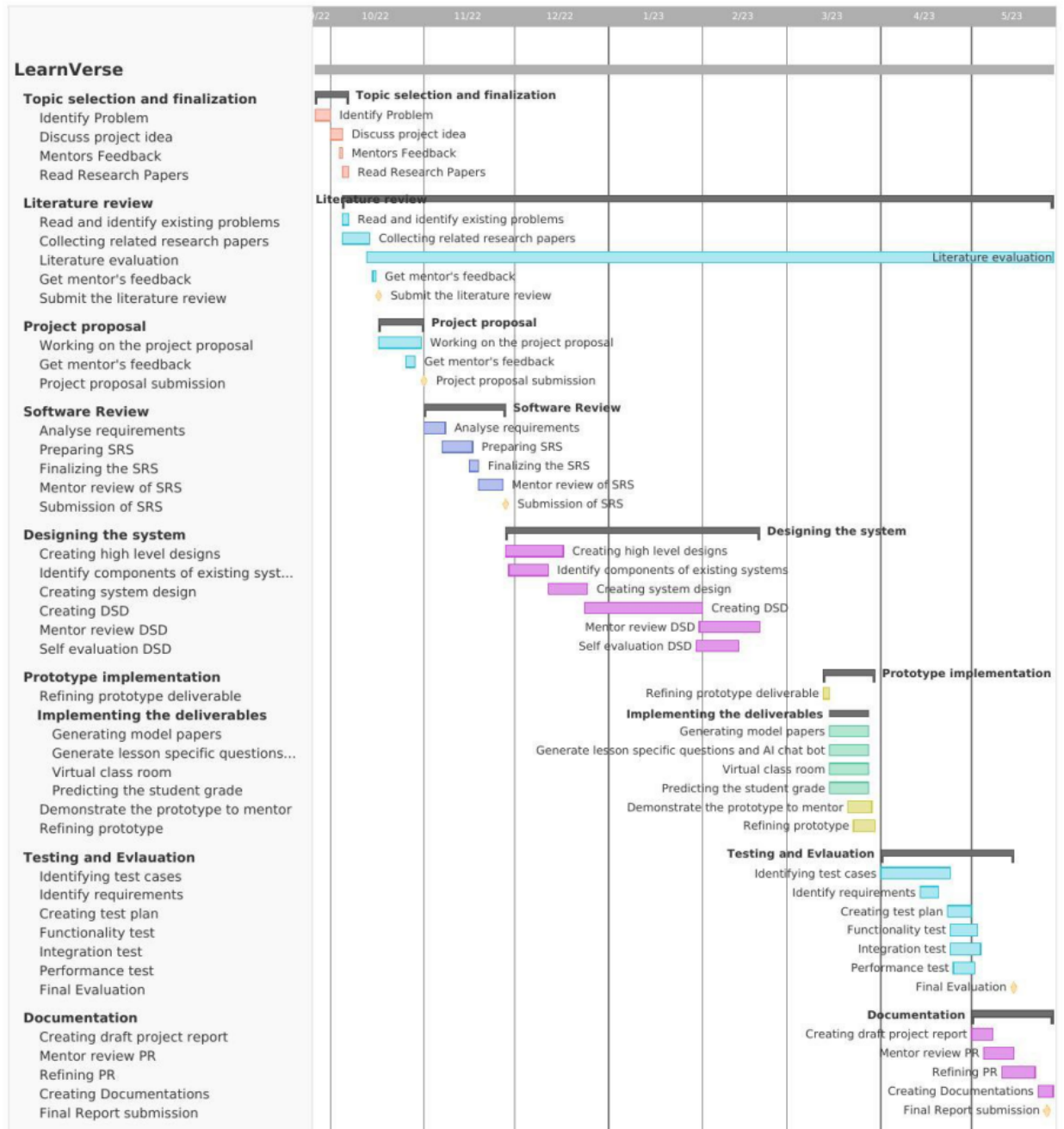


Figure 3 : Gantt Chart

## Deliverables

Deliverable	Date
<b>Semester 01</b>	
Submission of literature review	Week 3
Submission of the Project Proposal to the supervisor	Week 4
Submission of the Project Proposal(Final PP)	Week 5
Submission of the SRS to the supervisor	Week 8
Submission of the SRS (Final SRS)	Week 9
<b>Semester 02</b>	
Prototype Implementation	Week 14
Testing and Evaluation	Week 19
Documentation and final report submission	Week 23

## Reference

1. 3.5 Number Literate and Literacy Rates by Census years & Sex. (2015). [online] Available at: <http://www.statistics.gov.lk/Pocket%20Book/chap13.pdf>.
2. Wu, Z., He, T., Mao, C. and Huang, C. (2020). Exam paper generation based on performance prediction of student group. *Information Sciences*, [online] 532, pp.72–90. doi:10.1016/j.ins.2020.04.043. Available at: <https://www.sciencedirect.com/science/article/pii/S0020025520303716>
3. Vachev, K., Hardalov, M., Karadzhov, G., Georgiev, G., Koychev, I. and Nakov, P. (2022). Leaf: Multiple-Choice Question Generation. *Lecture Notes in Computer Science*, pp.321–328. doi:10.1007/978-3-030-99739-7\_41. Available at: [https://link.springer.com/chapter/10.1007/978-3-030-99739-7\\_41](https://link.springer.com/chapter/10.1007/978-3-030-99739-7_41)
4. Xu, B. and Liao, Y. (2021). *The Online Exam System Research Based on the MVC Framework*. [online] IEEE Xplore. doi:10.1109/IMCEC51613.2021.9482049. Available at: <https://ieeexplore.ieee.org/abstract/document/9482049>
5. Rahim, T.N.T.A., Aziz, Z.A., Rauf, R.H.A. and Shamsudin, N. (2017). *Automated exam question generator using genetic algorithm*. [online] IEEE Xplore. doi:10.1109/IC3e.2017.8409231. Available at: <https://ieeexplore.ieee.org/abstract/document/8409231>
6. www.sciencedirect.com. (n.d.). *Computers & Education | Journal | ScienceDirect.com by Elsevier*. [online] Available at: <https://www.sciencedirect.com/journal/computers-and-education>.
7. QuestionPro. (2019). *Evaluation Research: Definition, Methods and Examples*. [online] Available at: <https://www.questionpro.com/blog/evaluation-research-definition-methods-and-examples/#Quantitative-Methods> [Accessed 26 Oct. 2022].
8. Thomas, L. (2020). *Quasi-Experimental Design | Definition, Types & Examples*. [online] Scribbr. Available at: <https://www.scribbr.com/methodology/quasi-experimental-design/>.