PSYCHOMETRICS BASED RECOMMENDATION SYSTEM FOR HIGHER EDUCATION

Isuru Sapumal Perera

A dissertation submitted in partial fulfilment of the requirement for

Bachelor of Engineering (Honours) degree in Software Engineering

Department of Computing

Informatics Institute of Technology, Sri Lanka
in collaboration with
University of Westminster, UK

2020

DECLARATION

I declare that the work presented in this dissertation is my own work and to best of my knowledge acknowledgement is made for all sources of information used in this dissertation. Further, this as a whole or as parts has not been submitted previously or concurrently for a degree or any other qualifications at any University or Institutions of Higher Learning.

………………………………………….   …………..…………………...

        Signature of the Student   Date

Full Name of Student: Isuru Sapumal Perera

Student's Registration No: W1583345 / 2015562

The above Student carried out his /her research project under my supervision.

……………………………………   …………..…………………...

       Signature of the Supervisor   Date

Name of the Supervisor:

ABSTRACT

Higher education shapes the future of each individual and as well as an entire nation. Selecting an appropriate path to continue higher education is a crucial factor in the successful completion of higher studies. Today when a student steps into the higher education domain there are numerous decisions to make from the beginning itself such as selecting a proper course to follow from available choices. But due to the complexity of the current higher education domain, making these decisions are not easy.

Usually, at the point of starting higher education, most of the students are not fully aware of their preferences or potential of being successful in a particular pathway. So they are trying to seek supportive recommendations at this point. Even Though there are several existing mechanisms to make recommendations, those mechanisms are unable to match students' psychological behavioural factors with a success probability of a particular course in a quantitative way. There are many research-level approaches to predict student success in higher education. But most of these approaches have several limitations and gaps. Because of that even though these proposed mechanisms are helpful to achieve some research target or do some educational data mining they are unable to come out as a solution that can be used by any student and any higher education institute as a solution to the mentioned problem.

The aim of this research project is to identify the gaps between existing solutions and the actual requirement and build a feasible system to overcome those limitations. This system will have the feasibility to provide pedagogical support throughout the academic pathway which will not be limited to providing proper pathway recommendations but also can be used in many other related use cases. This proposed system will have the ability to adapt to the changes in curriculums without a system architectural or development change. This can be identified as an essential requirement when considering the long-term sustainability of such a system.

The author conducted this research project under the supervision of educational domain experts to ensure the required ethical and social aspects are considered. Several functional, architectural and algorithmic decisions were made during this research to make sure that outcomes of this research is feasible to achieve the research aim.

Please note that the proposed solution will be called "Plato".

Keywords: Decision Making in Higher Education, Recommendation System, Machine Learning, Psychometrics, Pedagogy, Personalized Recommendation, ML Automation

ACKNOWLEDGEMENT

I would like to take this opportunity to thank everyone who contributed to make project successfully. Completion of the project would not be possible if those who were with me could not give their advice, help and support.

Firstly, I would like to express my gratitude to Mr. Sameer Hamza, my supervisor and mentor for his support, guidance, and supervision throughout the project.

Next, I gratefully thank every lecturer, student, domain, and industry expert who dedicated their time to learn, use, and provide feedback on the project.

Finally, my heartfelt gratitude goes to my parents, my sister, Dilani Dassanayake and all the loved ones for the tremendous support and for always being there for me no matter what.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

Abbreviation	Definition
HE	Higher Education
ML	Machine Learning
IQ	Intelligence Quotient
ANN	Artificial Neural Network
DT	Decision Tree
GA	Genetic Algorithm
HT	Hyperparameter Tuning
MS	Model Selection
PT	Psychometric Test
RF	Random Forest
SVM	Support Vector Machine
NN	Nearest Neighbor
MBTI	Myers Briggs Type Indicator
PP	Performance Prediction
RS	Recommendation System

CHAPTER 1: INTRODUCTION

1.1 Chapter Overview

The purpose of this chapter is to provide an overview of the project structure, deliverables and details about the overall workflow of the research project. This chapter begins with a brief introduction about the higher education domain and the importance of selecting a suitable pathway with the problems faced by students in the process of making decisions. Also, this chapter presents an overview of the current student guiding methods with related problems. Next, describes the motivation of conducting this research followed by the research question, aim and objectives, and the contribution by the author. After that, the project scope and project deliverables are explained.

1.2 Problem Background

The education domain can be identified as a broad area of research. A large number of studies have been conducted in this domain from different perspectives. Innovations in technological applications and systems provide a crucial contribution to improving the quality of education. E-learning methods, distance learning, online courses, advanced learning management systems are being developed as a result of the combination of education and the growing technology. As a result of this revolution, the quality of education improves day by day by providing a more effective experience for the user.

After finishing secondary education, students tend to start HE which leads to the award of an academic degree. Many research can be found under the HE category to identify patterns and different approaches to enhance the effectiveness and the efficiency of the HE systems. Among them, automated applications for student guidance and providing recommendations are important to enhance the quality of the students' decisions. These researches make a vast value addition since, HE has a unique role to play in helping to achieve sustainable development goals and it equips students with the requisite knowledge, skills, values and attitudes to create a sustainable future. Moreover, HE effectively links with social and economic challenges in local, national, continental and global contexts.

Nowadays, researchers focus on implementing software-based to provide assistance in learning pathways to promote the learning performance of individual learners. As per that, personalised guidance and recommendations for students become a more significant topic which provides an adaptive learning experience for different learners by analysing individuals' skills, knowledge and learning preferences (Altun, 2012).

In order to provide personalised e-learning environments, there should be

- Learning Analytics - Tools for regular data monitoring and analysis.
- Learner's Characteristics - Determining of cognitive and non-cognitive characteristics.
- Learning and Instructions - Learner's interaction with designed mediums
- Ontology and Designing Navigational Paths - Tools for identifying or guiding learners with study or navigation paths (Altun, 2012)

Previous studies claim about factors that influence student success in a particular HE pathway (i.e. HE curriculum). Among them, cognitive and non-cognitive characteristics have a strong influence on students' performance as individual difference variables (Altun, 2012; Gray et al., 2014). The field of evaluating these cognitive and non-cognitive abilities in a person is well advanced these days and can be identified as psychometrics (Gray et al., 2014). To design a personalised recommendation system, it is important to guide students towards education that consistent with their psychometric (Altun, 2012) such as personality, cognitive abilities, preferences because those variables play an especially important role in identifying individual differences. At the same time, those differences can be used to classify students with similar traits.

1.3 Problem Domain

First of all, when it comes to selecting a pathway to continue HE, it is crucial to make the right decision by considering all available choices and other factors which can be found or extracted. Once a student passes out from a school, he or she has to make some crucial decisions regarding HE. Few amounts of selected students will have the chance to follow the degree programs offered by public universities. Remaining students have the ability to join a university or institute according to the expenses they can afford. Most of the time in

the real world at this step, students have to reconsider their goals, targeted jobs and rely on a possible and more realistic pathway to achieve it. After the affordable university selection step, the most crucial decision which needs to be taken by a student is selecting an appropriate, most suitable course to follow.

Accordingly, throughout the academic pathway, a student has to face many decisive points which require guidance and recommendations by considering his or her personality and preferences. Usually, many students struggle with these points. The most significant influence for the decisions on HE path selection is made by elder siblings, family status and school experiences (Ceja, 2006). The main reason for taking recommendations is they believe a wrong decision may waste their skills, abilities and ruin their future productivity even though they might not identify their skills and abilities accurately at this stage.

Taking a wrong decision can lead students to serious trouble when they are following the path which is not matching their preferences.

- Unsatisfaction of the student
- Low grades for modules
- Repeating modules (refer or defer)
- Can be a reason for early dropouts from the course (Najimi et al., 2013).

Also, these problems can be a cause for several other issues such as rework and repeating modules increase the cost of education. That may impact students directly and indirectly since it causes to improve the overall education cost of the country (Xu, Xing and van der Schaar, 2016). Another significant negative effect of that is it takes another student's opportunity who might have performed better in that pathway. As the above points clearly emphasis, it is critical to follow necessary steps and consider essential information before selecting a course, diploma or a degree program.

Furthermore, most of the applicants have to rely on traditional aptitude tests, and in case a particular test is not capable of evaluating the student's capabilities, personality and comparing those with course requirements, students will be appointed to a course or choose a module, which has no passion. Among the different methods students can use as a support for their decision-making process, as previously mentioned, understanding the course

content and doing a self-analysis about student's skills, preferences, and other psychological needs are important. At this phase, most of the students may search for recommendations and help from various sources to make an accurate decision. There are several organisations to provide guidance for students at different levels. Also, an option available to have a discussion with an educational counsellor to fulfil this requirement.

1.4 Problem Statement

Current student guidance mechanisms such as taking a counselling session can be comparatively expensive and can be dependent on the personal beliefs and understanding of the counsellor. Furthermore, available online guiding platforms are limited to evaluate some psychometrics from theoretical approaches such as knowledge-based systems and are unable to match with course requirements quantitatively. In the other hand, the course details and requirements which are provided by the university (i.e. HE Institute) is a broad set of information and not targeted to each individual separately. Such factors emphasis the requirement of a proper university-based RS to aid students to make useful and accurate decisions throughout the academic journey. Nevertheless, a RS should have the capability to adapt to changes in the educational domain without consuming more money and resources since the HE industry is changing rapidly by adding or removing contents in a curriculum. It is important to design a proper student guidance system which can be used as a supportive tool to make recommendations and suggestions along in the academic pathway.

1.5 Research Motivation

Education shapes the future of every person. In the 21st century, it is very important to get a proper HE to be a successful citizen. With the personal interest in the education domain, the author tries to find important gaps to fill in the current education system. One of the identified issues is there is no proper system to guide students to make more efficient decisions before selecting a HE path and also throughout their study pathway considering measurable psychometric variables. Experiencing these difficulties personally and also as a volunteer counsellor in an international organisation motivates the author to conduct research on design and develop a more feasible student recommendation system to assist students in making effective decisions throughout the HE journey.

1.6 Research Gap

Most of the Existing methods for providing recommendations in HE have several drawbacks and limitations. Suppose you consider student counselling that consumes more time and money. In addition to that qualified education, counsellors are not available in most of the areas. Nevertheless, this approach is unable to provide a quantitative success factor against a particular curriculum.

Taking a general online test to identify psychometrics and suitable recommendations only provide primary theoretical finding according to the existing theories. That approach doesn't validate student performance against a particular course or module or any specific factor. On the other hand, most of these solutions are for general career guidance.

Even though there are many research approaches to predict student performance using previous students' data, most of these researches contain considerable drawbacks such as,

- Drawbacks in student data gathering process which is not sustainable for long-term
- Designs to predict existing students pass or fail factor for a particular course
- Can only use predefined models to data mining.
- Only able to extract limited information from the predicting process.
- Only consider a particular scenario when designing the architecture which is not capable of fulfilling the different requirements in the academic pathway.
- No proper user interfaces or enterprise-level solution is delivered.

Furthermore, most of the proposed designs for university-based student performance are dependent on their specific environments and not targeted in commercial, economic aspects to make it a usable product.

1.7 Research Contribution

The Plato student guidance system is designed with the intention of avoiding previously mentioned limitations, barriers and providing a novel architecture to student recommendations systems. Firstly Plato contains a general test with unchangeable, general factors which can be used to compare students by measuring both cognitive and non-cognitive skills. Furthermore, by providing features such as detailed automated reports about recommendations, the ability to get an idea about the accuracy of the

recommendations to help both students and advisors to make more effective decisions. Nevertheless, it's designed as a highly controllable and adaptive system to maintain and upscale easily. Administrator dashboard of the system provides the facility to control many essential functionalities in terms of system sustainability.

Through proper research in behaviours of the machine learning algorithms, the Plato system includes optimised functionalities, automated prediction model selections, tuning, training and testing processes using various algorithms. Finally, this Plato architecture proposes solutions for some challenges that researchers are facing such as data limitations in the HE domain, what are possible areas to enhance a system in terms of providing recommendation etc. in future enhancements.

1.8 Research Questions

RQ 1: How to design and develop a maintainable and scalable student recommendation system which can provide reliable guidance throughout the academic pathway?

1.9 Research Aim

This research aims to design and develop a system which is capable of helping students by providing recommendations to make more accurate and useful decisions throughout the academic pathway with the ability to meet the industrial and domain requirements as a proper web-based solution.

1.10 Objectives

1.10.1 Research Objectives

ID	Objective Description
1	To identify the research and feature gaps in current student guidance mechanisms, systems and solutions • Assess the background of the research and the problem domain, which is guiding and providing recommendations in HE. • Evaluate research projects with similar goals and understand the different approaches, principles, different technologies and architectures • Identify the gaps between existing solutions and domain requirements.

2	To analyse the requirements for a proper system which is capable of providing recommendations to make effective decisions through the HE pathway. • Gather and analyse the requirements of possible end-users. • Acquire the insights of domain experts such as educational psychologists and counsellors for an effective solution. • Finalise the proper set of functional and non-functional requirements for the prototype of the proposing solution.
3	To develop a system which is capable of filling existing gaps and fulfil the requirements of possible end-users in the HE domain. • Design proper system architecture and workflows for a RS which should be capable of handling and analysing student data • Design and develop necessary algorithms and logics. • Develop the designed system according to the industrial requirements.
4	To evaluate the developed prototype to make sure the system is capable of meeting the requirements of end-users, domain experts and industry. • Test functionalities and the quality of the prototype by conducting a complete test against the functional and non-functional requirements of the system. • Evaluate the system by demonstrating the prototype to domain experts in education, ML and software developing industry.

Table 1. 1: Research Objectives

1.10.2 Operational Objectives

Objective	Description
Selection of the Project	To identify a gap in the particular domain and create a proper research question to propose a solution
Selection of the Methodology	To determine proper methodologies to conduct a research

Project Initiation Document	To prepare a PID which describes Problem domain, research gap, project scope, contribution, resource requirement etc.
System Requirement Specification	To create an SRS based on findings of requirement gathering phase of the research
Design Specification	To document the architectural and behavioural designs
Project Interim Report	To measure the research progress and document the status
Final dissertation	To document all the factors and attributes of the research project according to academic standards.

Table 1. 2: Operational Objectives

1.11 Scope of Project

Identifying the depth and scope of the research project is an essential task in the project planning phase. The defined scope should be realistic and achievable within the specified period using available resources. The initial scope should be wide enough to deliver a working prototype which has the capability to fulfil the research objectives in a satisfactory level. When defining the scope of the Plato system, more focus is given to provide an end to end solution more than increasing the number of choices available.

In-Scope

1. Classify students to the most suitable courses using psychometric data.
2. Predict student marks for particular modules and provide comparison reports
3. Provide detailed feedback with the recommendation
4. Provide extracted feedbacks for upcoming modules
5. Recommend educational resources based on psychometric data
6. Administrator dashboard to control the ML tasks.
7. Auto-select best ML model for the dataset, perform and deploy an optimized model
8. Manage student data sets

Out-Scope
1. Initially, the system will be developed to suggest a path between a minimum numbers of choices due to the data limitation.
2. Initial stage system will not use an abstractive summarizer in a student feedback recommending system.
3. Plato psychometric test will not contain essay type questions in this phase.

1.12 Existing Work

System	Ft1	Ft2	Ft3	Ft4	Ft5	Notes
Career Potli	AVL	P	NA	General	NA	Evaluate test results using rule based system
Mentoria	AVL	P	NA	General	NA	Evaluate test results using rule based system and suggest 3 study/career path in generally
mapmytalent.in	AVL	C	AVL	General	NA	Evaluate test results using rule based system and suggest a course from limited available courses.
careersportal.ie	AVL	P	NA	General	NA	Evaluate test results using rule based system and suggest 3 study/career path in generally

Table 1. 3: Hardware Requirements

Ft1: Consider Psychometric Result analyzing
Ft2: Suggest pathway P, Courses C, Module M, Resources R
Ft3: Consider course/module in an institute (specialized over courses in an institute-wise)
Ft4: Recommend a Specific/ General pathway
Ft5: Can improve over time using Feedbacks (i.e.: performances of the students who are currently following that particular courses)
Keys - *AVL - available * NA- Not Available

1.13 Resource Requirement

1.13.1 Hardware Requirement

• 16 GB Memory or above • 10 GB Storage or above, • Core i5 processor or above	To develop the Plato system (Includes train and test several ML/DL models, auto tuning hyper parameter that consumes considerable memory and processing power)

Table 1. 4: Hardware Requirements

1.13.2 Software Requirement

Software Requirement	Solutions
Languages (Programming, Scripting)	**Java, Python, Javascript** Base languages of the project
Frameworks	**Flask, Angular, Spring Boot,** Design and develop client-server architecture
Diagram Design Tools	**Creatly.com, draw.io** To draw diagrams and flowcharts for documentation
Project Management Tools	**Trello** For project management related works
Documentation Tools	**Microsoft Word, Microsoft Excel, Google Sheets** To document project-related work
IDE	**Jupyter notebook, IDEA, PyCharm, WebStorm** Integrated environments for programming
Packages and Libraries	**Scikit learn, pandas, Scientific Python, Matplotlib, Seaborn, Teapot AutoML, Tensorflow** For data mining and visualisation tasks
Languages (Programming, Scripting)	**Java, Python, Javascript** Base languages of the project

Table 1. 5: Software Requirements

1.13.3 Data Requirement

Reliable IQ , Personality ,EQ, Brain dominance tests	To create Plato psychometric test
Student marks (or range value)	Use as labels in ML dataset

Table 1. 6: Data Requirement

1.14 How User will benefit with the proposed solution

Plato Personalized student guidance system will provide direct benefits to university

candidates, students, student advisors and university management by providing

- Study pathway recommendation using various prediction, classification models
- Guidance for selecting modules and having an idea about upcoming modules
- Recommending more suitable educational resources
- Easy and useful features to manage and enhance system capabilities.

1.15 High-level Rich Picture
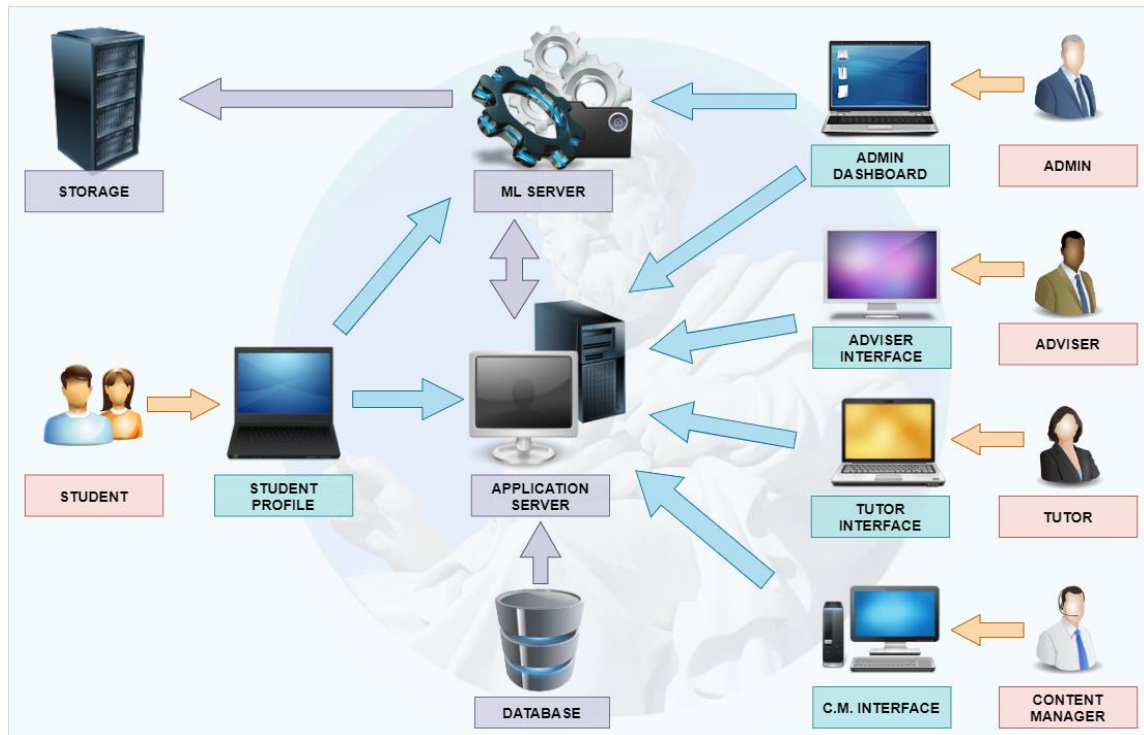
Figure 1. 1: Rich Picture diagram of the Plato System

1.16 Project Document Structure

This research project dissertation contains the following ten chapters. References, Bibliographies and Appendices are presented at the end.

Chapter1: Introduction chapter starts with the project background, problem domain, and problem statement of the project. Research motivation, gap, contribution, research questions to be answered, research aim and objectives, the draft of the literature, scope of the project are discussed. Further hardware, software requirements and project deliverables are documented.

Chapter 2: Literature Review chapter in the document provides analysis and discussions on the existing work and systems related to the project under three main categories in terms of domain, technology and existing products.

Chapter 3: Methodologies chapter gives an insight on methodologies followed during the research and development phases of the product. Selected research and software development methodologies and justification for the selections are included.

Chapter 4: Project Management chapter discusses the followed project management methodology and activity schedule, work breakdown structure, risk assessment of the project. Further presents how project compliance with the BCS Code of Conduct.

Chapter 5: SRS chapter covers requirements gathering procedures followed in the project. Stakeholder analysis, different methods of gathering requirements, use case diagrams, functional and non-functional requirements are documented.

Chapter 6: Design chapter of the document presents the design element of the system with appropriate diagrams such as high-level architecture, class diagram, context, activity, sequence diagrams of the system and explanations are provided.

Chapter 7: Implementation chapter describes how each aspect of the design was implemented. Selected tools, technologies, algorithms, core functionalities, development decisions and limitations problems faced during the prototype development are discussed.

Chapter 8: Testing chapter describes the testing procedure of the developed prototype. Both functional and non-functional testing of the prototype features, gathered test results are analysed and documented.

Chapter 9: Evaluation chapter covers how the project was evaluated to determine the success and impact. Selected evaluators, the reason for selections, received evaluations, authors evaluations are discussed and documented.

Chapter 10: Conclusion chapter presents the concluding remark for this research project. The author concludes the research project by discussing the retrospection on aim, objectives, limitations of the system, learning outcomes, and future enhancements of the prototype and compliance of legal, ethical, professional social aspects.

1.17 Chapter Summary

This chapter presented the decisions taken in the project initiation phase. The section mainly focused on discussing the problem background, problem domain, problem statement, research gap, research motivation, research contribution of the project, next the research question, research aim and related work were listed. Further, drafted the project scope, resource requirements and deliverables which will be required to the successful completion of this research project. Finally summarised the chapters included to this project dissertation.

CHAPTER 2: LITERATURE REVIEW

2.1 Chapter Overview

This chapter is focused on a critical review of currently published literature. The First section presents approaches of student guidance in HE domain, education psychology, and psychometric variable measuring mechanisms. Next, discussions about the technological approach in educational data mining, ML algorithm analysis and other related technology analysis will be covered in terms of predicting student performance and providing recommendations are included. Addition, the chapter includes analysis of existing work, commercially available products related in the domains

2.2 Concept Graph

Refer **Appendix B 1** for Concept Graph.

To achieve the research objectives of the Plato project, the author aimed to make and achieve below key points through this literature review

- The research background and importance of the correct decision making in HE.
- Popular mechanisms that students are using in decision making.
- Existing products for student guidance and recommendation in HE domain.
- The correlation between psychometric variables and student performance
- Reliable ways to measure psychometric variables
- The Importance of Educational data mining
- Various ML algorithms and their pros and cons
- The supportive technologies to achieve the research aim
- Existing researches in student performance predictions and provide recommendations

2.3 Literature Review of the Domain

2.3.1 Importance of HE and Selection of Correct Academic Pathways

HE, also called territory education or post-secondary education, is effectively engaged with the social and economic challenges in local, national, continental and global contexts (Badat, 2009). HE directly impacts the workforce of society because living status, salary

and reputation have been linked with the level of HE. Since HE has many positive impacts on a nation and its citizens, to seize successful achievements people should identify the appropriate HE path and follow that. Currently, few amounts of selected students have the chance to follow the degree programs offered by the public universities. The rest have the ability to join private universities and higher educational institutions (HEIs) and get the opportunity to select a pathway from many different diplomas and courses available. Even though economic factors are affected on the school leavers HE pathways selection, financial aid programs like student loans facilities, work-study programs have been introduced by universities and HEIs to minimise impacts from economic factors (Stiglitz et al., 2000).

The choice-based credit system is an advanced mode which allows students to have freedom in selecting their own choices for completing a degree program (Biswas, 2018) and it consists of 3 types of the main course categorisation as core courses, elective (optional) courses and foundation courses. Many students tend to pursue HE programs directly after school and do not have a clearly defined plan, because career exploration is limited in school life (Harrington and Orosz, 2018). High school advanced level subjects are practiced in different streams, it provides limited guidance for school leavers to select an academic path in HE. Usually, many students struggle at this point because wrong decisions may waste their skills, abilities and ruin their future productivity even though they don't identify their skills and abilities accurately.

Lack of success in selections leads to academic failures. Different kinds of aspects can be considered as HE failures such as repeating grades, dropping out, absence from classes and similar scenarios (Najimi et al., 2013).The problem of low academic achievement has become one of the significant issues in modern HE systems due to unavailability of proper guidance. Lack of success in HE leads to deviation from achieving future goals and it causes many social, mental and family problems for students. Research shows student loan debt in the United States has passed a trillion-dollar mark in 2014 and it is larger than auto loan debs and credit card loans combined. Also, students living expenses and tuition fees are getting higher day by day (Xu, Moon and Schaar, 2017). Abel and Deitz (2014) explain the two types of costs related to HE failures as increasing the direct cost for academic

expenses and the opportunity cost which occurs from the loss of the opportunity to earn income from a post-degree job. If people reduce the number of trials and retaking on a particular HE stream and achieve the qualifications within the minimum time possible, that helps to reduce the cost of HE and eventually it becomes more accessible for all, and the national success rate in HE will be increased. Likewise, it is essential to select an appropriate HE pathway, and next to its section critically evaluates approaches of HE pathway selections.

2.3.2 Evaluation on approaches for seeking opportunities in HE

As the first step to the HE, the selection of the right university or HE program becomes a critical and complicated task. There are several approaches that direct students to study pathway decisions throughout HE. As main approaches of the selection process student tend to select pathways based on specific career goals, through general information searching and academic counselling and guidance.

Discussion about mentioned approaches are listed in **Appendix - B2**.

2.3.3 Evaluation of student guidance and support techniques in HE

When it comes to selecting and guiding newcomers to universities and HE institutions, they review student portfolios using individual or general criteria. Some of the available techniques in HE support, guidance and students retention are listed in **Appendix- B3**. However, it has been proven that individual differences should be considered (Al-Zoubi and Younes, 2015) rather than giving guidance using general methods.

2.3.4 Review of psychometrics variables influence in academic performance

Previous studies and researches claim that most of the school leavers' decision on academic pathway selection depends on the influence of some other people or external facts. But it is vital to consider personal characteristics before deciding on an academic pathway. Al-Zoubi and Younes (2015) claimed that behavioural traits, psychological factors, intelligence are important in explaining achievements in HE because those factors are not easy to control and make differences in ability and students specific learning styles influence academic achievements. Psychometrics can be identified as theories and techniques of psychological variables (Hobart and Cano, 2009) which are concerned with studying differences between individuals.

Psychometrics are used as both cognitive and non-cognitive variables that have a strong influence on students' performance (Altun, 2012). Working capacity, perception, attention level, language skills can be considered as cognitive skills and in other hand personality, motivations, like, attitude, anxiety (Altun, 2012; Bjorklund-Young, 2016) are categorised under non-cognitive skills. When developing an educational environment effectively, it is important to reduce stress while maintaining students' psychological balance with wellbeing (Tosevski, Milovancevic and Gajic, 2010). Therefore literature review investigates how students' personality, EQ, IQ and brain dominance can be utilised as more elaborated variables to measure students' academic performance which can be implemented to provide useful recommendations in HE pathways.

2.3.4.1 Personality Types

Personality can be known as one of the widely identified and most critical indicators for individual differences (Irani et al., 2003). According to the classical definitions for the personality, it is described as an internal factor which gives consistency for the individual's behaviours (Tlili et al., 2016). Furthermore, Tlili et al. claim that the person's real-world behaviours, tastes, and actions are significantly associated with the personality, and this difference is used when making professional or personal decisions and judgments. Several past studies have provided answers to the question of why students' personality is an important factor in HE. Those studies explored the correlation and patterns between academic performance and personality traits using standard personality models like Myers-Briggs type indicators (MBTI) and Five-Factor Model (FFM) of personality.

2.3.4.1.1 Personality Types and Academic Performance

Chamorro-Premuzic and Furnham (2003) have investigated two British university samples over a period of three years in relation to academic performance(i.e., exams, final year projects), the big five personality and some academic indicators(i.e., essay writing, tutors exam predictors). They provide evidence related to personality traits such as conscientiousness type students are positively associated with academic performances as well as extraverts often tend to fail their courses than introverts due to impulsiveness and sociability. That research concludes that personality factors are a more significant predictor of exams, final year project grades and overall rather than other academic indicators and

tutors' exam predictions (Chamorro-Premuzic and Furnham, 2003). Numerous studies are shown that different personalities also affect subject-specific performances. Kun, Kiss and Kapitany (2015) have statistically analysed a sample of two reasonably similar subjects at the University of Debrecen. An example of 224 second and third-year students relevant for BA in Business Administration Management (BAM) and BSc in Business Informatics (BI) programs had been analysed using Myers-Briggs personality types and found ESFJ type students showed significantly better performance than ENFJ, INFP, ISFJ types in BAM program and ENFJ type students performed well. In comparison, ESFJ-type show middle and ESFP-type students produced worse performance in the BI program (Kun, Kiss and Kapitany, 2015). Based on studies conducted in the software engineering students and industrial software engineers, ISTJ and ESTJ were found as the majority and well-performed personality types and INTJ suitable for designers (Gulati et al., 2016).

Also, the preferred learning style varies depending on personality temperament (Fallan, 2006). Based on another statistical research using MBTI personalities Fallan (2006) explored that the SJ type students prefer to study economics and finance programs which include facts and procedures and NT, SP, NF temperaments prefer subjects that emphasis people and human relations such as health care, psychology and humanities. A better understanding of the student's personality is useful in appropriate academic, career choices and as a prevention for mental health problems which cause for most of the academic failures as well (Tosevski, Milovancevic and Gajic, 2010).

2.3.4.1.2 Modern Approaches to Measure Personality Type

Personality models have been developed to identify individuals' behaviours and characteristics. These models present different theories related to personality traits. Based on the defined methods, standardised personality scales and questionnaires have been developed to assess individual differences. Myers-Briggs type indicators (MBTI), Five-Factor Model (FFM) also known as Big Five Factors and Minnesota Multiphasic Personality Inventory (MMPI) can be identified as few of them. Comparison for the mentioned indicators is included in **Appendix B 4**

2.3.4.1.3 Myers-Briggs Personality Types Indicators (MBTI 16 Personalities)

The Author tended to select MBTI personality test questionnaires ethically for the Plato guidance system as it contains questions that are easily understood for the user with everyday experience. This test is extensively used and validated in the education domain (Tlili et al., 2016) as well as the widely accessible personality inventory more than 3.5 million worldwide tests (Sach, Petre and Sharp, 2010) is administered each year over sixty years to find the best fit type of personality. This indicator measures how people perceive the world and make decisions for mental preferences. Sixteen types of personalities have been defined along with four scales, and each has two opposing choices. These dichotomies have been listed based on

- Orientations of Energy - Extraversion (E) or Introversion (I)
- Preferred ways to gather Information - Sensing (S) or Intuition (N)
- Preferred methods to make Decisions - Thinking (T) or Feeling (F)
- Orientations towards dealing with external World - Judging (J) or Perceiving (P)

TABLE 1.   The four polarities describing a person's personality type

E: Extroversion	I: Introversion
The person's interest flows mainly to the outer world of actions, objects, and persons	The person's interest flows mainly to the inner world of concepts and ideas
S: Sensing	N: Intuition
The person prefers to perceive the immediate, real, practical facts of experience and life	The person prefers to perceive the possibilities, relationships, and meanings of experiences
T: Thinking	F: Feeling
The person prefers to make judgments and decisions objectively, impersonally, considering causes of events and where decisions may lead	The person prefers to make judgments and decisions subjectively and personally, weighing values of choices and how they matter to others
J: Judgement	P: Perception
The person prefers mostly to live in a decisive, planned, and orderly way, aiming to regulate and control events	The person prefers mostly to live in a spontaneous, flexible way, aiming to understand life and adapt to it

So

Figure 2. 1: MBTI personality types (Fallan, 2006)

Characteristics of mentioned temperament types can be identified as specified in Figure 2.1 MBTI 16 personality types Figure 2.2 are built up as a combination of the said personality preferences pairs.

Someone's trait is symbolised by using four letters which represent the person's temperament (Tlili et al., 2016) (i.e., ESTJ that stands for Extraversion, sensing, Thinking, and Judging personality). MBTI Company established in 2018, now the company provides training, guidance and certifications for the psychological counsellors to maintain well-being of the society.

ISTJ	ISFJ	INFJ	INTJ
ISTP	ISFP	INFP	INTP
ESTP	ESFP	ENFP	ENTP
ESTJ	ESFJ	ENFJ	ENTJ

Possible MBTI Types

Figure 2. 2: 16 Personalities (Sach, Petre and Sharp, 2010)

2.3.4.2 Intelligence Quotient (IQ)

Intelligence can be defined as the general mental capability that includes the ability to think, plan, solve problems, think abstractly, understand complex ideas, judge and comprehend, learn quickly and learn from experiences (Kiss, Kotsis and Kun, 2014; Bjorklund-Young, 2016). Intelligence is also known as cognitive skills. There are many ways to measure intelligence, the most popular being IQ tests of cognitive tests. Different types of cognitive abilities such as mathematical skills, verbal skills, reasoning skills, spatial skills can be measured using IQ tests (Carter, 2005). Aptitude tests are designed to measure students' intelligence in HE admissions (i.e., SAT-Scholastic Assessment Test, ACT-American College Testing).

2.3.4.2.1 IQ and Academic Performance

Researchers provide evidence that standardised aptitude test results positively correlate with academic performance in HE. Further evidence available for its ability to predict individuals' performance using intelligence (Komarraju, Ramsey and Rinella, 2013; Bjorklund-Young, 2016). Simultaneously, recent studies reveal that IQ alone is not a definitive predictor for academic performance as only 20% of success depends on the IQ (Mohzan, Hassan and Halil, 2013; Bjorklund-Young, 2016). Therefore, measuring only cognitive skills cannot be considered as an appropriate approach to guide individuals on a particular path. Both cognitive and non-cognitive skills need to be considered to guide someone and predict the future more efficiently (Altun, 2012, Bjorklund-Young, 2016).

2.3.4.2.2 Modern Approaches to Measure IQ

Various models and theories have been defined to measure the potential for cognitive skills. Among them Raven's Metrics, Raymond Cattell, Stanford- Binet and Wechsler intelligence test models describe different types of cognitive skills to be considered to

measure IQ levels of individuals using different scales, i.e., Wechsler test focuses on measuring verbal comprehension, working memory, perceptual reasoning and processing speed of a person (Mackintosh, 2011). Testing intelligence is a well-known approach of identifying and classifying newcomers in the admission processes in HE and job placements. Most of these tests are not based on intellectual theories but, they assess cognition based on reasoning, understanding, memorisation and problem-solving skills, etc. to make sure an individual's ability to perform well in HE or as an employee.

2.3.4.3 Emotional Intelligence (EI)

Emotional intelligence is defined as a subcategory of the social intelligence that involves "the ability to monitor one's own and others' feelings and emotions, to distinguish among them and to use this information to guide one's thinking and actions" (Mohzan, Hassan and Halil, 2013; Kiss, Kotsis and Kun, 2014). People use these skills to read, understand, and react effectively for the emotional signs of others and to themselves. Mayer and Salovey (1990) proposed the concept for emotional intelligence and Daniel Goleman (1995) defines major domains in EI as

- Self-awareness (Awareness of self-emotions, self-confidence and accuracy self-assessment)
- Self-management (Ability to control self-emotions and behaviours, adaptability, trustworthy, to be conscientious)
- Social-awareness (Being empathetic to others and having a service orientation)
- Social skills (Relationship management, demonstrating leadership skills, mentoring others, showing good teamwork, conflict management and collaboration skills) (MacCann et al., 2019).

2.3.4.3.1 Emotional Intelligence and Academic Performance

EQ is considered pivotal for a student's health and for success to avoid failures in academic environments. Researches have shown that students with emotional intelligence, such as a high level of self-awareness, motivation, and good teamwork, are more likely to cope with demands and complex experiences. When individuals are able to lead their personal and social life successfully, they can focus on their learning and academic achievement (Mohzan, Hassan and Halil, 2013).

2.3.4.3.2 Modern Approaches to Measure Emotional Quotient

As a significant non-cognitive skill, a person's genuine emotional intelligence level (or emotional quote) can be measured in the same way that cognitive skills are used. Ability-based EQ tests are designed to measure individuals' EI, and the results show whether higher or lower EQ for personal and social competencies. Similar to other skills, EI also can be acquired and improved with practice. Therefore EQ may change very little over the lifespan of a person based on the experiences.

2.3.4.4. Brain dominance

Left and right brain theory have been built based on the functional differences in left and right hemispheres. This concept was introduced in 1960 by Roger W. Sperry (Singh, 2015) who was a Nobel laureate in neurobiology. Studies have proven that the majority of the human brain is dominant, which means that one side of the brain, either the right brain or the left brain, functions well. The term brain dominance or brain lateralisation is known as the way a person wants to learn, understand and express something (Singh, 2015) depending on an approach of the one brain hemisphere is used more than the other. Roger explains that if a person is mostly logical, analytical in thinking, better at verbal, and computations that person is declared as left brain or digital brain. On the other hand, the right brain or analogue brain person is more intuitive, visual, and more creative (Genovese, 2005).

- The left brain is connected to make objective judgments, control feelings, intellectual, reliance on remembering, and prefer multiple choice answers tests with logical problem-solving, structured and planning.
- The right brain is connected to symbolic instructions, subjective judgments, prefer open-ended questions with intuitive problem-solving, and prefers to draw and manipulate objects.

As two sides of the brain perform differences in-process information, it is important to know about the individual differences as well. However, a small number of individuals possess the quality of bilaterally that both hemispheres are integrated when making decisions.

2.3.7.2 Brain Dominance and Academic Performance

Researchers explored that the brain hemisphere dominant can be used for individual matching and classify Genovese, 2005. Singh (2015) conducted a study with a 200 student sample and identified the difference between there behaviours.

2.3.7.3 Modern Approaches to Measure Brain Dominance

There are several ways to measure the brain dominance of a person. Questionnaire based approaches to find out whether the particular person is more logical or creative, Video or image based approaches to identify what are the dominant sides of the body can be identified as common approaches. Herrmann Brain Dominance Instrument (HBDI) recognize as a standard and valid measurement approach of the brain dominance. HBDI introduce a fourfold quadrant model of brain dominance. This instrument is widely using for scoring the brain dominance. Further, HBDI uses the MBTI concepts in the scoring process to measure variable such as introvert, extrovert characteristics as a part of its measuring process.

2.4 Literature Review of Technology
2.4.1 Educational Data Mining

Today, Most of the educational institutes manage large amounts of student data that is stored in their databases. With the rapid growth of new academic paths, curriculums and storing techniques, the amount of these records are increasing day by day. It is crucial to convert these massive amounts of data to useful knowledge which can be identified as meta-knowledge about education. This field identifies as educational data mining (Asif et al., 2017). EDM applies ML, statistics, data mining etc. on educational data to recognise patterns, information to make decisions (Dutt, Ismail and Herawan, 2017). (Baker and Yacef, 2009) proposed EDM can mainly contain one or more methods from the following 5 mechanisms which are predictions from data, recognize relationships, cluster data, Findings in models, and extraction of data for human judgment. Further explains the use of unique data for psychometric models, pedagogical supports. The Plato system uses a combination of the above techniques to achieve its goal of guiding students. EDM research mainly focuses on the following areas (Romero and Ventura, 2010) Effect of Psychological

factors on human learning patterns. This area is focused to identify the relationship between students' psychometric data/behaviour and performance in different curriculums.

- E-learning and LMS - Improve EL and LMS provide communication, collaboration tools and evaluate students in long-distance education.
- Intelligent tutoring systems, which are able to evaluate students.

Today EDM uses different ML techniques such as association, classification, regression and clustering to discover different types of knowledge (Baradwaj and Pal, 2011). The Plato system mainly focuses on the recognised relationship patterns between student's psychometric data and student performance and uses this knowledge to guide students through the academic pathway. Usually, various types of ML algorithms can be used to identify these relationships and patterns.

2.4.2 Algorithmic Analysis (Overview in machine learning)

2.4.2.1. Comparison of ML Algorithms

Machine learning is simply described as it lets computers have the ability to learn without human involvement for explicit programming. (Olson and Moore, 2016). There are various ML algorithms to achieve different goals in different environments. ML algorithms can be simply categorised by their learning mechanism and usages as follows.

Figure 2. 3: ML Algorithms

Supervised Learning	These algorithms train the model using labelled data. Model recognise patterns between independent variables (i.e. features) and dependent variables (i.e. labels) (Singh, Thakur and Sharma, 2016) Algorithms: DTs, Neural Networks, Linear Regression,
Unsupervised Learning	Can be used when data is not labelled These algorithms are capable of identifying similar features from data. ex: cluster unknown data set Algorithms: K-means Apriori, PCA
Semi-Supervised Learning	This mechanism can be used for a dataset with only a smaller amount of labelled data. Ex: Generative models, Self-Training
Reinforcement learning	RL is the area of ML that it learns from its own behaviour. RL can be used for the scenarios that any training data is unavailable. Its Behavior is vastly different from Supervised and unsupervised ML. Usually, task should has specific characteristics to apply reinforcement learning ● Agent ● Reward system ● Trial and error process/ states with proper instructions ● Suitable environment to proceed actions (Initially simulated environment) Algorithms: Monte-Carlo methods, Q learning.

Table 2. 1: Analysis of ML Algorithms

We can further divide different ML types according to the output it provides and these algorithms can be used in different scenarios and different phases in the ML life cycle.

Type	Description
Classification	Algorithms which can be used to identify the category of an object **Algorithms:** SVM, RF, Naive Bayes, DTs
Regression	This can be subdivided as logical regression, ordinal regression etc. These types of algorithms can be used to predict associated attributes. **Algorithms:** Ridge regression, Lasso, RF regression

Clustering	Auto grouping algorithms come under this. **Algorithms:** K-means, DBScan, Hierarchical clustering
Association	Association rule algorithms are used to identify the relationships and associations among the available data. Many areas such as Product recommendations, DB researchers, information science etc. using these algorithms to discover important associations **Algorithms:** Apriori
Dimensionality Reduction	Dimensionality reduction is an important step of the ML lifecycle in almost all the ML projects. DR use to reduce the number of dimensions (features) that we are using to resolve an ML use case. DR can be categorised in many ways such as feature extraction or feature selection, Supervised or unsupervised, linear or multiple. **Algorithms:** PCA, SelectKBest

Table 2. 2 Different usages of ML algorithms

Identifying proper boundaries among target classes and selecting the most appropriate target class for a set of features identified as classification in ML.

2.4.3 Algorithm comparison

2.4.3.1 Comparison in Deep learning vs Traditional ML Algorithms

Deep learning is a subset of ML and can be recognised as a ML method based on data learning features. Today DL has numerous applications in several domains (XIN et al., 2018). DL can be more efficient than traditional ML in areas like image or speech recognition, text generation, dealing with high dimensional datasets etc. ANNs, convolutional networks, recurrent neural networks can be identified as examples for DL algorithms. Nevertheless, there are some scenarios that Traditional ML algorithms are more suitable than DL algorithms. Below grid contains some cons of DL when compared with Traditional ML (Awad and Khanna, 2015)

Traditional ML Algorithms	**Deep Learning Algorithms**
Performs well in small data sets since it contains established rules in the algorithm.	Performance isn't efficient with small data sets. Requires large amounts of data for

	good accuracy. Performance increases rapidly with the amount of data available.
Domain knowledge can be used to feature engineering. The learning process can be compared with domain knowledge.	Features will be directly extracted from the data. No human involvement is necessary for this phase. Also, unable to justify the learning process using domain knowledge.
Consume comparatively less computational power	Consume more computational power. (Ex: GPUs)
Comparatively less execution time	Execution time will be high in most cases.
Interpretability is high	Interpretability is low

Table 2. 3: Comparison of Traditional ML Algorithms vs Deep learning

2.4.3.2 Genetic Algorithm

GAs can be identified as an advanced, effective selection algorithm which can find an optimum solution for a given population. These algorithms are mimicking the natural evaluation process. Usually, GA follows a few steps to find the optimal from a population. Those are briefly,

- Formulate population,
- Initialise population randomly.
- evaluate objectives in order to find a fitness function
- Apply genetic operators. Those Are
  - Reproduce,
  - Apply crossover
  - Mutate slightly until GA finds the optimum solution or meets the terminating criteria. (Kabra and Bichkar, 2014)

GA is useful in parameter optimisation in ML models. Since GA is commonly used in evolving ML. It is proved that using GA is an effective way ML parameter tuning (Kabra and Bichkar, 2014) Initially the ML model of this proposed system (Plato) will be contained attributes with various weights including i.e. If there is less or no evidence how

age or gender acts in ML algorithms initially their weight might be 0. With the increment of feedback data, GA can be used to reanalyze the weight and optimize the ML model by parameter tuning.

2.4.3.3 Classification algorithms

Refer **Appendix B 6** for **Classification Algorithms** comparison

2.4.3.4 Regression algorithms

Refer **Appendix B 7** for Regression Algorithms comparison

2.4.4. Introduction to Automated Machine Learning

Since the data preprocessing phase is completed MS and HT tasks are the most important phases in the ML life cycle. Usually, these steps are coming with a lot of trial and error. In traditional ML, it is important to find out the most suitable classifier, regressor or another type of algorithm to build the best possible model. Also, in DL, it is important to find out things like how many hidden layers should be applied for better results.

 In the HT phase, it tries to find out the best possible parameter values and combinations for better performance of the model. Usually techniques like GridsearchCV, RandomizedsearchCV are used here. Today there are Several AutoML algorithms to successfully automate the MS and HT part of the ML lifecycle. These AutoMl algorithms make it easy for domain specialist, data analysts to create an ML model without having in-depth knowledge of several ML algorithms. Below table evaluates some available AutoMl algorithms.

AutoML Library	Pros	Cons
TPOT Tree bases optimisation technology 1.Contains Regressor and classifier method 2.Use GAs to Find the best hyperparameters	Process a large number of possible pipelines before selection. Considered as the best autoML library for regression problems since it provides models with less MSE when comparing with other AutoML libraries (Balaji and Allen,2018)	Can consume more time to provide an optimised model since the GA based solutions are evolving with the time. Unable to process natural language inputs

Auto Sk learn build on top of Scikit learn Packages. Use Bayesian Search to optimise the model.	Easy to use. Many numbers of classifiers, feature & data preprocessing methods are available.	Not good for large data sets, Dependable on OS. Currently supports only for Linux
Auto Keras Built on top of Keras DL framework	Can use to tune DL models	Hard to get an idea about the selected model
ML box Support state of the art algorithms	Offering, Data cleaning, Feature selection, HT functionalities.	Dependable on OS. Currently supports only for Linux

Table 2. 4 AutoML Library Comparison
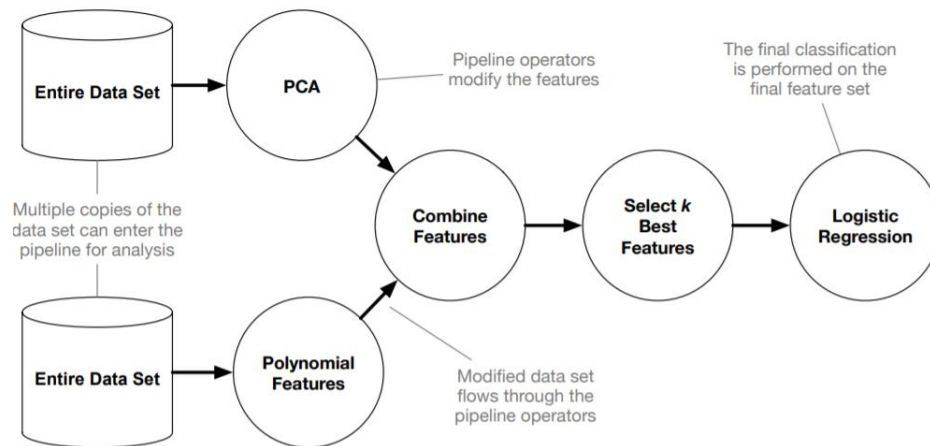
Figure 2. 4 TPOT AutoML (Feurer and Hutter, 2018)

2.4.4.1 Instruction and a brief overview of Federated Learning

This section briefly describes a federated learning concert in ML and its possible applications in the HE domain. When conducting this research a major problem that the author faces was the restrictions and limitations in gathering required student data. As a

possible future solution for that this federated learning concept was identified by the author during the research. Currently, there are several use cases for this concept mainly with CNN and RNN. Even Though applying this for a domain like education is beyond the scope of this research, these findings are documented for possible future enhancements of this project.

What is Federated Learning.

It is vital to deal with an enormous amount of data when training an ML model. But privacy issues can occur in the data collecting processes. Therefore it might not be possible to centralise all the data to a particular server and train the model. When it comes to sensitive user data such as health information, military data it is not ethical to acquire that data and make it accessible for a third party. Even if data can be anonymous up to some extent, still there might be a situation that reveals information. Hence authorities, organisation, users might not be agreeing with collecting the data for a third party. As a solution to this concept called federated learning is introduced. It is a distributed ML approach (Bonawitz et al., 2019).

How Federated Learning Works

In Federated learning technique, the ML model will be deployed to the user's (participant) machine (PC, Laptop or mobile device). The model will train inside the device using user data for a limited number of iterations within a certain time period. Then the updated model will be uploaded to the central server. As such, a central server will aggregate all the contributions and create a new version of the model by comparing all. Then this updated version of the model will be redistributed to the users to train further. When the central server identifies an optimum model, this process will be terminated (Yang, Liu, Chen and Tong, 2019).

What is the importance of Federated learning in the education domain?

Student data such as personal details, module marks, psychometric data, and recorded behavioural patterns of students can be categorised under confidential or sensitive data. Most of the universities and HE institute keeps this data under high security to make sure

third parties have no access. But when it comes to applying data mining and ML mechanisms to these data, it always prefers massive data sets to obtain better accuracy and performance in these tasks. Therefore federated learning can be used to train these models with the data from multiple higher educational institutes for better results without compromising students' privacy.

2.5 Literature Review of Existing Systems

Using ML Algorithms with Past Performance and General Information to Predict performance

Among various applications of ML, classification and regression-based applications are widely used in many domains. ML has been popular in the education domain for the past few years. There are many studies that investigated the ways of using ML for various educational purposes. (Agrawal and Mavani, 2015). Student's performance prediction is one of them and its efficiency has rapid growth. Various researches proved the efficiency of ML to predict performance using different types of parameters. Also, there are various types of ML algorithms which were used in performance prediction such as DTs, SVMs, and Neural Networks etc. Furthermore, much research proves most of these ML approaches show considerably high accuracy rates in prediction tasks (Pal and Bhatt, 2019). Almost all these researches have proved that increasing data set size may increase the accuracy rate of prediction.

It is crucial to critically evaluate their approaches and results they obtained by using various types of data and different prediction algorithms. Bhardwaj and Pal (2011) explored the patterns and relationships in a student's academic performance and predicted future performance using Bayesian classification techniques in data mining. As an initial dataset, they collect the records of 300 students. The Importance of this research is it tries to build connectivity between various non- academic factors including student's financial income, family background, food habits etc. and academic success. Researchers provide possible predefined categories for the above factors and use those to gather information in a quantitative format. Above research concludes that there are various other factors which can affect a student's performance other than curriculum knowledge and grades. Even

Though there are many factors which can affect a student's performance, those factors can be different from area to area or country to country. Therefore it is more important to find common factors and build a proper model using those to predict a student's academic performance. Using factors like student brain dominance, personality type and other related psychological data to train the model is reliable since those factors will be common for each and every student.

Hamsa, Indiradevi and Kizhakkethottam (2016) conducted research by predicting student performance using DTs and fuzzy GAs (FGA) separately. One objective of this research was to compare the performance and the accuracy of DTs and FGA algorithms. As a training data set, they used past academic marks of students. Results of this research say FGA categorized more students as safe performing and DTs were more biased to categorize students under a risk category. This gives an idea about the algorithmic behaviour of different algorithms on the same student dataset.

Agrawal and Mavani (2015) published a research paper to propose a model to predict student academic performance using ANNs which is a broadly used ML technique. In this research, they also try to compare bayesian classification against ANN and prove ANNs will perform with better accuracy when the data set is comparatively large. In most of the approaches, the initial dataset will be small. Therefore the accuracy of ANN classification might be lower. Yao et al. (2019) published a detailed research paper on predicting students' behaviours, performance. Mainly researchers collected two categories of data. Academic performance data and behavioural data. In order to predict behaviours, they have collected many types of data by observing students and tracking their digital student identity cards rather than giving a questionnaire to the student and deriving information through that. In the behavioral data set, it includes sleeping patterns, time spent in the library, data on breakfast and other meals etc.

However, the important part of this research is it tried to consider the big five personality traits to their behavioural prediction. In order to analyze these personalities, they extracted two types of predictors as orderliness and diligence.asl behavioural related data was categorized on these two and analyzed. Researchers declared a few simple algorithms to

express the connectivity between considered factors. Also above research validates that if two selected students have similar behaviors then they may have similar academic performance. They collected university student data from 2011 to 2015 and used them as a training data set for their proposed algorithms. In order to train the model, from a ML perspective they used several algorithms such as DTs, RF to predict performance and Rank SVM to compare methods. The research concludes by saying there are significant correlations between student behaviors, characteristics and academic performance. Also, they highlight that there can be various other factors which can affect students' performance and they didn't consider them due to the limitation of data collection.

Halde, Deshpande and Mahajan (2016) conducted some similar research by using Neural Network and DT algorithms together to predict performance. In their data-gathering phase, they provide a questionnaire consisting of 98 questions which contain questions about past performance, LASSI Scales and personality traits (Halde, Deshpande and Mahajan, 2016). Later they trained the ML model with psychological factors and without psychological factors. Results of this research prove the accuracy of the prediction was high when the model uses psychological factors with other data instead of ignoring psychological factors. Above research is a good source to prove that there are various other factors which can affect students' academic success and those factors can be presented in a quantitative format. Since the researchers manually gather a lot of behavioural data by observing students for a long time this approach may not be good with newcomers or fresher because it is not possible to collect all information in a short time period. Even though this approach needs long term data it's a trustworthy way of having an idea about university students' future performance as a whole. Also, an important factor of above research was they show that not only previous academic qualifications but also various other factors such as living location, the medium of teaching etc. can be affected to the student's academic performance. Bhila (2018) conducted research to emphasize the importance of a system to manage students by helping them. Even Though this research is conducted for primary and secondary schools she highlights that this kind of system will be a good platform for all students, teachers, parents and policymakers to have a logical idea about the efficiency of the particular education process.

Summarized comparisons on proposed performance prediction systems

Gray et al, 2014 published a review paper on existing research articles regarding student performance prediction with psychometric variables. They summarize the usages of classification algorithms in PP, what was the intention (i.e., prediction class label) number of data sets and obtained accuracy. Further it provides an idea about the relationship between number of classes, size of the data sets and the accuracy of the prediction models.

Study	Algorithm	Accur-acy	n	Class label	VLE	ITS
Baker et al. (2011)	Bayesian Network (BKT)	AUC: 0.7029	76	next question correct		x
Merceron & Yacef (2005)	Decision Tree (C4.5)	87%	224	pass/fail		x
Minaei-Bidgoli et al. (2003)	Ensemble	94%	227	pass/fail		x
Pardos et al. (2011)	Ensemble (Neural Networks)	AUC:0.77	5,422	Performance on ITS		x
Romero et al. (2008)	Fuzzy Rule (MaxLogit-Boost)	62%	438	module performance 4 bins	x	

n=number of instances; AUC=Area under the Curve

Figure 2. 5: The usages of classification algorithms in PP (Gray et al, 2014)

In the same paper they reviewed some classification approaches in PP using different types of data. As they emphasize even though the number of instances are less psychometric data provides considerable accuracy.

Study	Algorithm	Accur-acy	n	Class label	VLE	ITS
Baker et al. (2011)	Bayesian Network (BKT)	AUC: 0.7029	76	next question correct		x
Merceron & Yacef (2005)	Decision Tree (C4.5)	87%	224	pass/fail		x
Minaei-Bidgoli et al. (2003)	Ensemble	94%	227	pass/fail		x
Pardos et al. (2011)	Ensemble (Neural Networks)	AUC:0.77	5,422	Performance on ITS		x
Romero et al. (2008)	Fuzzy Rule (MaxLogit-Boost)	62%	438	module performance 4 bins	x	

n=number of instances; AUC=Area under the Curve

Figure 2. 6: Classification approaches in PP using different types of data (Gray et al, 2014)

Gaps between aforementioned prediction models/ systems and domain requirement

- Target to achieve some research objectives rather than designing a commercially viable solution
- Limitations of the scope and the available features.
- Lack of self-sustainability as a system.
- Lack of extensibility, scalability, maintainability, security in proposed systems

2.6 Chapter Summary

This chapter focused on reviewing facts which can affect student performance in HE, critically evaluate existing mechanisms in HE for student guidance and provide recommendations. Thoroughly discussed how personality types, IQ, EQ and brain dominance as measurable, behavioural, intrapersonal attributes affect student performance. Selected Then algorithmic review is conducted on ML algorithms to compare the efficiency and accuracy. Further existing researches on student performance prediction and recommendation were discussed. Finally, a brief introduction about federated learning and a description of how it may help with this kind of a system.

CHAPTER 3: METHODOLOGY

3.1 Chapter Overview

This Chapter includes methodologies followed in different phases of the life cycle. The chapter initiates by discussing research methodologies and why they are suited for carrying out the project. Next, it will discuss the selected software development methodologies for successful completion of the project.

3.2 Research Methodology Approach

Research methodology can be known as studying how the research will be carried out. It is essential to consider suitability, the accuracy of the result and efficiency for the selection of methodologies for a research-based project (Rajasekar, Philominathan and Chinnathambi, 2013). Research onion model (Saunders, 2009) used to identify the suitable methodology. Table 3.1 represents the selected research methodologies and justifications for the selection.

Layer	Selected Method	Justification
Research Philosophy (Paradigm)	Constructivism	Research paradigm selection is important to gain reliable knowledge of the research objectives. Among positivism, social constructivism and mix mode philosophies **Social constructivism (Qualitative)** was selected because open-ended questionnaires will be used to collect data and knowledge for a social phenomenon (Hyde, 2012) through the perception of social actors and their counterproductive actions as there is no existing model for benchmarking this research.
Research Approach (Execution Mode)	Deductive	Among the deductive, inductive and mix approaches, **deductive** was chosen as the most suitable approach based on the research questions and the data needed for solving those questions. As the qualitative approach provides an understanding of the phenomenon in the

		context, the research proposition formed using the pre-existing theories and validated published instruments. The aim of this research is to analyse and confirm the proposition using data gatherings and testing. Hyde (2012) explains qualitative research frequently adopt deductive process something very like hypothesis testing.
Research Strategy	Interviews Surveys	Appropriate research strategies are important in order to address research questions. Among experiments, Surveys, case studies, action-oriented researches, interviews etc. **interviews and surveys** were selected to gather data, opinions and suggestions. Under those structured and unstructured interviews, questionnaires, discussions, brainstorming and observations used to improve the research.
Research Choice	Mix Method	Choice of research interacts with methods to gather types of information. Among mono, mixed and multi-methods, **mixed-method** was chosen because it involves collecting, analysing and integration combination of both qualitative and quantitative methods to create a precise set of data. Qualitative methods include interview responses, observations and brainstorming. Survey questionnaires are used to find information on attitudes and behaviours as quantitative data gathering methods.
Time Horizon	Cross-sectional	Time horizon deals with the required time for data gathering for the project completion. Among the cross-sectional and longitudinal horizons **cross-sectional** was chosen since the data gathering will be done once during the research by observing the behaviour of a group of people.

Data Collection and Data Analysis Methods	Techniques and Procedures	This layer consists of techniques and procedures used for data collection and analysis. Based on reliability and validity, primary data will be collected through interviews and surveys. Interviews will be conducted with few psychological and higher educational domain professional experts and surveys will be conducted with the target audience. Secondary data will be gathered from, existing literature in related domains, organisational documents.

Table 3. 1: Research Methodologies

3.3 Software Development Methodology

Software development methodology can be known as a recommended means to achieve the development of a system based on an underlying philosophy and set of rational (Vavpotic and Bajec, 2009). It includes definitions for the development phases, procedures, tasks, tools, techniques, rules, guidelines and documentation for a project. Characteristics, pros and cons of some software development methodologies are included in **Appendix-C1**.

Independent variables like project size, development style, changes of requirements, development time and risk need to be considered when selecting a software development methodology (Öztürk, 2013). Among the Agile (Scrum), RAD, Spiral, Waterfall, Prototype development methodologies, the **Spiral** software development methodology was selected after considering the nature of the project. The Plato project is developed as a medium scale project. When considering the development time, though this project should be delivered during the 7 month period, the spiral methodology supports continuity of the software development life cycle and allows to add additional functionalities in later, using pre-defined prototypes

Relevant factors to the Plato System	W	A	S	R	P
Project size and complexity - Plato research project is a medium scale complex project which involves a single developer and no large number of stakeholder involvement in the development stage.	✖	✔	✔	✖	✖
Changes in requirements - Requirement fluctuations are applicable for this research project based on the survey and the interview data gatherings, the methodology should be flexible for changes	✖	✔	✔	✔	✔
Development style-Iterative and incremental methods are suitable for this research project as products developed on defined prototypes	✖	✔	✔	✔	✔
Development time-Major functionalities should be developed during the 7 months of the period in advance but the methodology should be supported for the continuous refinement of the final software product based on pre-defined prototypes.	✔	✖	✔	✖	✔
Risk management - As the project domain is complex and the ML (Federated learning) ideas are new to the author a considerable amount (Medium) of risks are involved in the project.	✖	✖	✔	✖	✔

Table 3. 2: Analysis for most suitable software development methodology for Plato system

Keys – W-Waterfall, A-Agile, S-Spiral, R-RAD, P-Prototyping, ✔- Suitable ✖ - Not Suitable,

Further, this methodology focuses on an advanced risk assessment process. Risk factors related to the new development technologies and project domain are considered and avoidance of risk is involved in every phase. Due to the time constraints and frequently changing requirements, Spiral methodology will be the best fit software development methodology for the development of this research project.

3.4 Chapter Summary

This chapter started with justifying the selected research methodology for the project using the research onion model. Qualitative method selected as the research philosophy and deductive mode was chosen as the research approach according to the nature of the project. Next, analysis and design models were discussed. The spiral model was selected as the best fit software development methodology for the project.

CHAPTER 4: PROJECT MANAGEMENT

4.1 Chapter Overview

The previous chapter discussed the research methodology, data gathering, software development, design methodologies of the research project. This chapter includes the project management methodologies followed during the project. The chapter starts by identifying appropriate project management methodology and next moves on time management and resource requirement of the project. It further discusses the identified possible risks and mitigation plans. Finally, constraints and dependencies will be discussed as challenges for the project.

4.2 Project Management Methodology

Project Management Institute (2020) explains project management is the application of skills, knowledge, tools and techniques to the project activities to fulfil project requirements. This process falls into groups as initiation, planning, execution and closing (Larson and Gray, 2020). It is essential to manage project scope, time, cost, quality, resources, integration and risks in a methodical approach from the initiation to finishing the project life cycle. PRINCE2, PMBOK, CPM are some identified methodologies in modern project management. **PRINCE2** (Projects IN Controlled Environments) methodology was chosen as the project management methodology because it provides manageable, controllable stages to develop and deliver the project. With the PRINCE2 method, the project outputs are clearly defined and delivered on time. This methodology includes 7 processes to control and organise a project more effectively.

- Directing a Project
- Start-Up a Project
- Initiation of the project
- Controlling a Stage
- Managing a Stage Boundary
- Managing Product Delivery
- Closing a Project

When compared to other methodologies these steps are easy to understand the involvement of project management to the project and easy ensure the project viability document is updating is critical at every project stages, project is planned and controlled stage by stage, estimations can give by the lessons learnt from the previous experiences in software

industrial work, the adaptively to project environment (with risks, importance, time) and a quality of deliverables should be ensured as defined. Therefore this methodology was chosen as the most suitable project management Methodology in this project lifecycle.

4.3 Tasks and Time Allocation

To complete the Plato project within the defined deadline Work Breakdown Structure, Gantt chart used to identify main tasks and allocated time for each task. These techniques make managing work more effective and easier. Trello tool is used as the project management tool in the project.

4.3.2 Activity Schedule

Table 4.1 represents an overview of the expected main tasks of the project and time durations according to the deadlines of the main deliverables. The literature review process for the research projects to be carried out till the requirements specification or test phase due to frequent requirement variations although the draft submission due early. Subsequently, the project documentation will be carried out throughout the lifetime of the project as it is the final deliverable.

	Task Name	Planned Start Date	Actual Start Date	Planned End Date	Actual End Date	Duration
1	Draft Project Initiation Document	4-Aug-19	4-Aug-19	23-Sep-19	23-Sep-19	50 Days
2	Prepare Final Project Initiation Document	23-Sep-19	23-Sep-19	20-Oct-19	20-Oct-19	27 Days
3	Literature Review	10-Aug-19	10-Aug-19	14-Nov-19	14-Nov-19	96 Days
4	System Requirement Gathering	20-Nov-19	20-Nov-19	14-Jan-20	14-Jan-20	55 Days
5	System Designing Phase	30-Nov-19	30-Nov-19	14-Jan-20	14-Jan-20	45 Days
6	Prepare and Submit Project Interim Report	20-Jan-20	20-Jan-20	10-Feb-20	10-Feb-20	21 Days
7	Development and Implementation	25-Jan-20	25-Jan-20	30-Mar-20	30-Mar-20	83 Days

8	Testing	28-Feb-20	28-Feb-20	20-Apr-20	20-Jun-20	72 Days
9	Evaluation of System	2-Apr-20	30-May-20	17-Apr-20	20-Jul-20	50 Days
10	Final Project Report	10-Aug-19	10-Aug-19	30-Apr-20	27-Ju-20	352 Days

Table 4. 1: Activity Schedule

Please refer **Appendix – D2** for detailed Work Breakdown Structure.

4.3.1 Work Breakdown Structure

Please refer **Appendix – D1** for Work Breakdown Structure.

4.3.3 Gantt Chart

Please refer **Appendix – D3** for Gantt chart of the Project.

4.4 Resource management

Resource management should be proceeded alongside the project management to make sure the best utilization of resources in achieving research requirements. When considering human resource management, since this project should be carried out as an individual research project as per the requirement, all the roles in the project life cycle should be played by the author. Hence the author fulfils the project manager, business analyst, developer, tester roles throughout the project life cycle. The supervisor monitors the status of all the phases and progress of the work in a timely manner. Other resources such as finance should be allocated as required for ensuring a quality output. Buying licenses for software and tools, domain name registration and gathering important literature which are unavailable for free is required suitable finance allocation.

4.5 Risk Assessment and Mitigation Plans

This section describes the risks identified during the project life cycle which will block the development process. Project management, Technical, Resource risks were identified. Risk assessment was done using risk probability and impact and possible mitigations plans are included. Used 1-5 range numeric scale to define the impact and probability of risk assessment.

Risk Type	Risk Factor	Impact	Probability	Mitigation Plan
Project Management	**Limitation of time** Time can be a very challenging fact due to tight deadlines of submissions because some phases will be required more time than expected.	5 High	2 Unlikely	Prepare a time plan for each phase and analyse activity schedule Meet supervisor and updating the progress
Technical Risk	**Unavailability of data** Open data sets are not available with suitable variables to train ML code and it will take considerable time to gather a new data set. This may cause delays in achieving deadlines in the activity schedule.	5 High	3 Medium	It should be searched and reassured the availability of open data sets. If still, data is not available the author needs to collect a data set to train and analyse the ML model.
Technical Risk	**Failure to keep up to date** ML technology and FL, domains are very new, and there were researches conducted related to the domain and similar products may launch after the project initiation	3 Medium	2 Unlikely	Check domain area and technology changes on a weekly basis, communicate with domain experts to know about the latest development. Analyse impact if new competitors are made,
Resource Risk	**Failures in data and software resources** During the project	5 High	2 Unlikely	Maintain a daily backup procedure, continue

	development process, it can get lost data due to Software failure.			debugging the software and using version control in software like GitLab.
Resource Risk	**Failures in Hardware resources** The memory of the hardware may not be sufficient.	3 Medium	3 Medium	Test the system before start working

Table 4. 2: Risk Assessment and Mitigation Plans

4.6 Special Challenges

4.6.1 Constraints and Dependencies

Constraint 1	Time Constraint
Time considered as a major constraint	
Constraint 2	**Lack of prior Knowledge in new technologies**
Lack of knowledge on ML related developments will be affected to spend more time on the development and testing phases of the Project.	
Constraint 3	**A large corpus of Decentralized data sets**
Limited time-frame will be a constrain to collect large amounts of decentralized (university-wise) datasets to implement Federal Learning Module	

Table 4. 3: Constraints and Dependencies

4.6.2. Compliance with BCS, Code of Conduct (SLEP analysis)

According to the standards of this undergraduate degree program, as a Software Engineer undergraduate it is essential to conduct this research project according to the formulated rules and professional standards by British Computer Society (BCS). "Profession should have due regard for public health, security, privacy, wellbeing of others and the environment" (BCS, THE CHARTERED INSTITUTE FOR IT CODE OF CONDUCT FOR BCS MEMBERS, 2020). The outcomes of the projects should be benchmarked with the BCS code of conduct. This section includes the author's own judgment to meet

requirements of legal, ethical, social and professional aspects of the overall project using SLEP method.

S- Social Aspects
- Self-Improving Student Guidance System will be developed with social interest to make positive value addition to the educational domain   - The project ensures to enhance the decision-making process and make life easier with a more personalised suggestion to select study pathways. (i.e., Plato system lets students consider the possible success factor when making decisions and related guidance.)   - Project contribution will be directly involved to increase the quality of the decisions taken by students, tutors and advisers.   - All the libraries, resources and tools used for this project were properly licensed or open-source hence no third party rights violation.   - No religion, nationality, race, gender appropriations are applicable to the project.
L - Legal Aspects
- The data collected from the Plato system will only be used to train the system.   - Student module marks will be provided by students anonymously.   - Any of the data collected for the project will not be legally affected by any party.   - Personality type test (MBTI), IQ test, EQ test and brain dominance tests and validation processes were selected from the standard publications (Online, Printed) which were published by domain experts.
E - Ethical Aspects
- The Plato system will only require the personality type, IQ, EQ and brain dominance and module marks of the students.   - Above features were selected after proper research on the educational domain and also after considering marketing aspects. Therefore the Plato system will not collect any sensitive private information.   - At the initial training stage of the ML model, the data will be collected fully anonymously.   - The proposed architecture of the Plato system itself is designed to protect the students' privacy and achieve maximum efficiency (using Federated Learning module, Different access levels to the system)

• Evidence from literary works for this project has been properly cited. • Questionnaires respondents were informed about the project and how their contribution to the project will be handled.
P - Professional Aspects
• The Plato system introduces a novel approach to guide students to make more accurate decisions throughout their academic pathway. • The final product of this solution will be provided as a readymade system and it will be customized to the specific requirements of the particular higher educational institute. • All the control of the system and access levels will be maintained by the particular higher educational institute. • This research project maintains professional integrity in all areas in line with the BCS Code of Conduct.

Table 4. 4: BCS, Code of Conduct (SLEP analysis)

4.7 Chapter Summary

This chapter initiated by explaining the importance of selecting a project management methodology for the project. Among the widely used project management methodologies PRINCE2 was selected as the most suitable methodology due to the flexibility and developer has to less involvement for the learning of project management procedure. Project work was broken down into an activity schedule according to the main deliverables of the project. To show the time management, work breakdown structure and Gantt chart were created for time allocation in project work. Next resource requirement of the project was represented. Identified risks to the project listed in the risk assessment section and mitigation plans were discussed. Communication plans and project monitoring plans with the project supervisor were explained in the project supervision section. Some constraints and dependencies were identified as special challenges to the project. Finally, SLEP analysis was done to identify social, legal, ethical and professional aspects which should be completed within the project as compliance of BCS code conduct.

CHAPTER 5: REQUIREMENT SPECIFICATION

5.1 Chapter Overview

This chapter focuses on gathering and analysing the requirements of the proposed system. Under the overall description, the first section describes associated stakeholders and the specific roles related to the system. Requirement elicitation techniques section will provide methodologies used to gather requirements and outcomes after executing them. Moreover Context diagrams, use case diagrams and descriptions of the proposed system are included in this chapter. Finally, external interface requirements, functional and nonfunctional requirements are described.

5.2 Stakeholder Analysis

Stakeholder engagement with the system is visualised using onion diagrams. Stakeholder roles and identified relationships with the system are explained

5.2.1 Onion Model

Figure 5. 1: Onion Diagram

5.2.2 Stakeholder Roles and Viewpoint Analysis

It is essential to identify the role and viewpoints of stakeholders who are interacting with the system in-detail, the table 4.1 figures roles and viewpoints of each identified main stakeholder for the Plato system.

Stakeholder	Role	Viewpoint
Students	Functional beneficiary	Students are able to get psychometric base recommendations when making academic selections throughout their academic life.
Academic Staff	Operational Maintenance	Engages to enter data such as students marks and feedback related to student performance
System Administrators	Operational Administration	Directly engages with Plato maintains and opens financial model by operating the system
Academic Advisors	Functional Beneficiary	Use the system as a helping tool to identify and classify students to provide guidance.
University Management	Financial Beneficiary	Easier to manage decisions in admissions (course selections) with less time and resources consumed than doing it manually and easier to classify students for courses and acquire financial benefits, when students' success rates increase through Plato, it increases the reputation of the university and increases opportunities for new comers.
Sponsors		Sponsors receive a percentage of the profits made after the establishment of the Plato at a HE centers.
Product Owner		Manages constraints and dependencies to ensure the smooth flow of the work product and provides the developer with the functionality needed to maintain the system.
Product Developer		Creates the proposed system that meets requirements and get financial benefits when developing a bug free system.

System Security Professionals		Experts for safety as they know where system security is at risk and attacks may occur.
Parents of Students	Functional beneficiary	They finance academic activities, and the success of students is important to them.
Competitors	Negative Stakeholders	Can build a system with similar features
Hackers		Hack and interrupt the system.
Educational Psychologists	Domain Experts	Provide specialised expert opinion on the methods and techniques used for the project.
Public	Usability Opinion	Benefit from getting feedback on software usage and pointing out weaknesses in the solution

Table 5. 1: Stakeholder Roles and viewpoints

5.3 Analysis of Requirement Elicitation Techniques

Requirement elicitation is the process of gathering all requirements of the system from stakeholders. Several elicitation methods and approaches were explored from various stakeholders to ensure the effectiveness of the requirement elicitation process. Chosen methods, advantages, disadvantages and the summary of the findings are discussed.

Method 1	Observing Existing Solution in the Market
As the initial step of the utilised methods in requirement elicitation was observing existing student guidance systems. With this domain analysis, identified feature gaps and helped come up with new requirements on the advanced student guidance system.	
Advantages	1. Identified the importance of a recommendation (guidance) system 2. Identified limitations of the available student guidance systems
Disadvantages	1. Time consuming 2. Required to go through several publications and websites in detail

Findings: Most available RSs are designed to provide some general overview of suitable pathways to follow. Those systems don't provide any quantitative recommendation for any specific courses available in a university. Also, most of these systems require human experts to provide personalized guidance.

Method 2	Literature Review

An appropriate literature review was important to gather requirements on domain information like existing systems similar to the project and technologies, research approaches with similar goals and related work which belong to the HE recommendations systems

Advantages	1. Get a proper understanding about the existing system and limitations of technologies relevant in existing work. 2. Widely accepted recently updated information were gathered
Disadvantages	1. Time-consuming as it requires more time and effort to review a large number of literature for an extensive analysis in the domain.

Findings: As a significant contributor to this research project, literature review findings supported to gain a proper understanding of the domain, theories and technological approaches and further requirements related to the system. Repositories like ScienceDirect, IEEE Xplore, Springer, were used to find literature. Findings are documented in the Literature Review chapter.

Method 3	Interviews

Interviews with domain experts and related specialists are crucial to validate the requirements. Hence found and interviewed some domain experts to make sure that the research is heading to the right direction and also to identify necessary adjustments in requirements. Since this project is related to the education domain, psychology and ML. **Dr. Kanthi Hettigoda** the Sri Lankan President of the International Union of Psychological Science (IUPsyS) (Psychologist in Sri Lanka Navy and Lecturer) and **Ms. Sithari Tilakaratna** (Educational Counselor in University of Colombo Faculty of Education) were interviewed to identify problems in current student counselling methods and gathered further requirements to develop the system. Also **Mr. Lahiru Jayakody** (Software Engineer - ML Researcher) was interviewed to have an idea about ML requirements in such a research.

Advantages	1. Easy way to get clarifications on doubts 2. Identified problems related to psychometric factors and satisfaction level of students could be easily discussed
Disadvantages	1. This method cannot be used to reach a large expert audience due to the time schedule needed to carry out interviews. 2. Gathered requirements during the interview may not be a problem faced by the entire students in society.

Findings: Based on Dr. Kanthi Hettigoda's explanation, psychometrics such as personality type, IQ, brain dominance are important when making decisions in HE. In addition, the psychologist advised the author to consider the Emotional Intelligence (EQ) level of students for the Plato system because EQ also has a significant effect on individuals' behaviours and success. Educational psychologists mentioned how educational counselors provide guidance for pathway selections and about techniques of how they identify the personality of students

Method 4	Brainstorming

The author conducted several brainstorm sessions to validate requirements.

Advantages	Understand the project from different perspectives to present new features and development mechanisms
Disadvantages	Some conflicts happened with the ideas shared in several discussions

Findings: A number of discussions were carried out as brainstorming sessions. Project supervisor, seniors in previous batches and colleagues were part of these sessions. Problems with designing the system, discussing problems, algorithms, the flow of the project were solved with the contribution of brainstorming.

Method 5	Self-Evaluation

Many self-evaluation sessions were conducted by the author.

Advantages	Ability to identify new requirements, boundaries that have not been identified specialists

Disadvantages	Submission to self-experience may overlook some important requirements

This was really helpful to make lots of decisions and identified possible enhancements.

Method 6	**Questionnaire Distribution**

In order to collect data and identify the requirements of the end-users, questionnaires were prepared in Google Forms and distributed through emails and social media platforms as an online survey. This was one of a simple, efficient, and cost-effective method of gathering data. The design of the questionnaires is included in the **Appendix E-1 section.**

Advantage	1. It was able to cover a larger range of global audiences which included the target audience and geographically available users. 2. This saves time compared to the other methods 3. Result analysing was easier with functionalities in Google Forms tool.
Disadvantages	As this process is limited to standardised questions, users may be restricted to sharing additional information.

Table 5. 2: Review of Requirement Elicitation Methods

5.3.1 Questionnaire Findings

The author distributed a questionnaire via email to the general public (2020-01-12). Another questionnaire (2020-04-08) was distributed among IIT students to gather their opinion on the skills required for different types of modules. With the aim of achieving the main objectives of the research project, questionnaire related objectives Table 5.3 were identified for the survey because the information needed to generate the specifications required for the project can be effectively grabbed.

5.3.1.1 Operationalisation Table
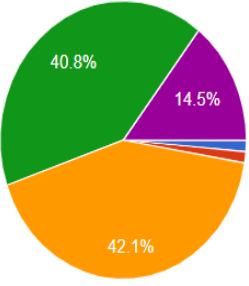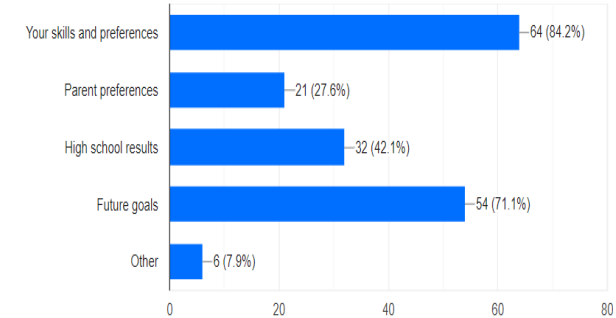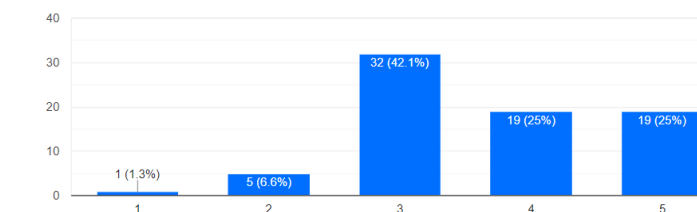
Objective	Question No
To identify answered audience belongs to research of proposed system	1, 2
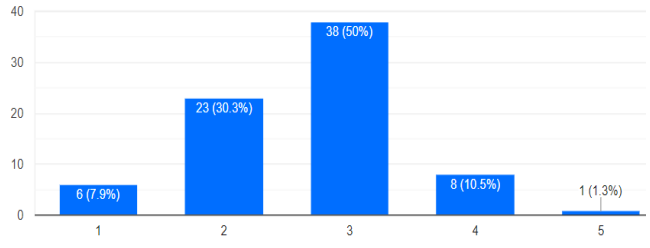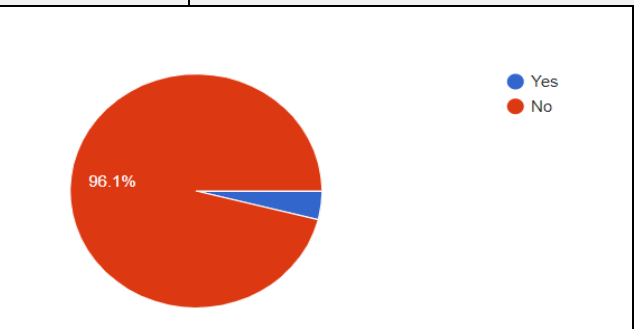
To identify the gap between available mechanisms, existing products and end-user requirements	3, 5, 6, 7
To determine and the prioritise the required features of the system	4,8
To validate the overall user requirement of the proposed system	9

Table 5. 3: Operationalisation Table

Questions are formulated based on their relevance to different functional areas, results and the conclusions of each question are discussed.

Question 1	Who are you?
Question 2	What is your education status?
Results	85.5% of responses were received from students, and 11. 8% of the audience were employees. As a majority, 92.1% of the audience was undergraduates.
Conclusion	Aimed to identify the answered audience and educational status which they are belongs to. The questions were used to filter out the respondents who are suitable to respond to the survey. 2.6% of respondents were school leavers, while the other total of 97.4% had experienced selections in HE as undergraduates, graduated, and MBA students. Hence all are valid respondents.
Question 3	**Have you ever faced any difficulties when making a decision about academic pathways?**

	Results 40.8% of the respondents answered that they have often faced difficulties, 42.1% of the audience responded as sometimes, and 14.5 % said always, and, 1.3% of the audience said they rarely faced difficulties. Another 1.3% said they have never faced any difficulty when making decisions on an academic path
Conclusions	Aimed to ensuring that the audience faces the problem related to the research. Based on the answers received from respondents' experiences it was validated as the audience practically face for the problems in academic selections.
Question 4	**What do you mainly consider when selecting a degree or diploma?**
	Results Majority of the audience consider approaches like skills, preferences and future goals. Some are tend to select pathways based on high school results and parents preference. Only 7.9% focuses beyond.
Conclusions	Aimed to prioritise approaches to select academic pathways. It was able to identify that candidates are considering their own skills and preferences and future goals when they are selecting a pathway.
Question 5	**Are you satisfied with the chosen academic path to follow? (1-extremely dissatisfied, 5 extremely satisfied)**
	Results 42.1% of the audience presented a mixed opinion as majority while 25% has shown they are extremely satisfied with the chosen pathway.

Conclusions	This question is focused on finding the current satisfaction level of the audience with the chosen academic path based on previous mentioned approaches. Whether they have selected the correct path or not. Result proved that there are some gaps should be address to increase the effectiveness of the HE selections.
Question 6	**Are you satisfied with the current student guidance mechanisms and recommendation methods? (1-extremely dissatisfied, 5 extremely satisfied)**

Result

50% of the audience expressed mixed feelings while the majority of the others were dissatisfied.

Conclusion	Focused to analyse the current satisfaction level of users and the gap between available approaches and technologies to student guidance. It can be clearly seen that the users are unhappy with the available mechanisms. More factor should be consider when academic pathway selection than general approaches.
Question 7	**Have you heard about any automated student guidance system or any course recommendation system?**

Result

96.1% of the respondents have not experienced HE recommendations with modern technologies. 3.9% had responded they have experienced, but the answers were not related to automated systems.

Conclusion	Objective was to identify technical approach availability related to the purposed system and to do a further observation of available HE recommendation systems. To verify whether there is any competitor who could not be analysed during the existing systems observations and literature reviews. But the majority proved that they have not experienced such.

Question 8	According to your view, Is it important to consider the personality type, IQ and brain dominance before choosing a course or module?
	Result 64.5 % of the audience agreed that the psychometrics like personality, IQ, and brain domains should be considered in HE pathways selections. And 31.6% responds with mix feelings and only 3.9% are not aware on those factors.
Conclusion	Aim was to determine and prioritise the required features for the system. More than the half of audience validated the psychometrics like personality, IQ, and brain domains should be considered in HE pathways selections.
Question 9	When you are making a decision about your academic path, Would you like to receive a suggestion from an automated system by analysing your personality type, skills, and IQ?
	Result 85.5% of the audience agreed with such a solution and expected to get benefits from the solution in this project. 11.8% have responded with mixed feelings. Only 2.6% have thought that kind of solution is useless to select the most suitable academic path to follow.
Conclusion	With the objective of validating the overall user requirement of the proposed system was verified as that the system will be added value for the target audience. The majority of the responses were encouraged to continue research on the suggested project as the target audience is expecting to have such a solution to avoid difficulties faced in academic selections.

Table 5. 4: Questionnaire Findings Analysis

5.4 Summary of Findings

Keys: **Q**-Questionnaire, **B-**Brainstorming, **I-**Interviews, **S-** Self-Evaluation, **L-** Literature Review

No	Findings	Q	B	I	S	L
1	Should be capable of recommending a module to a student by analyzing psychometric tests and predicting their success factor	✔			✔	✔
2	The system admin should be able change the prediction and it should be easy to configure		✔			✔
3	Should be able to view the recommended feedback gave for previous students with similar traits			✔	✔	✔
4	Should use EQ test for Plato psychometric test			✔	✔	✔

Table 5. 5: Summary of Findings

5.5 Context Diagram
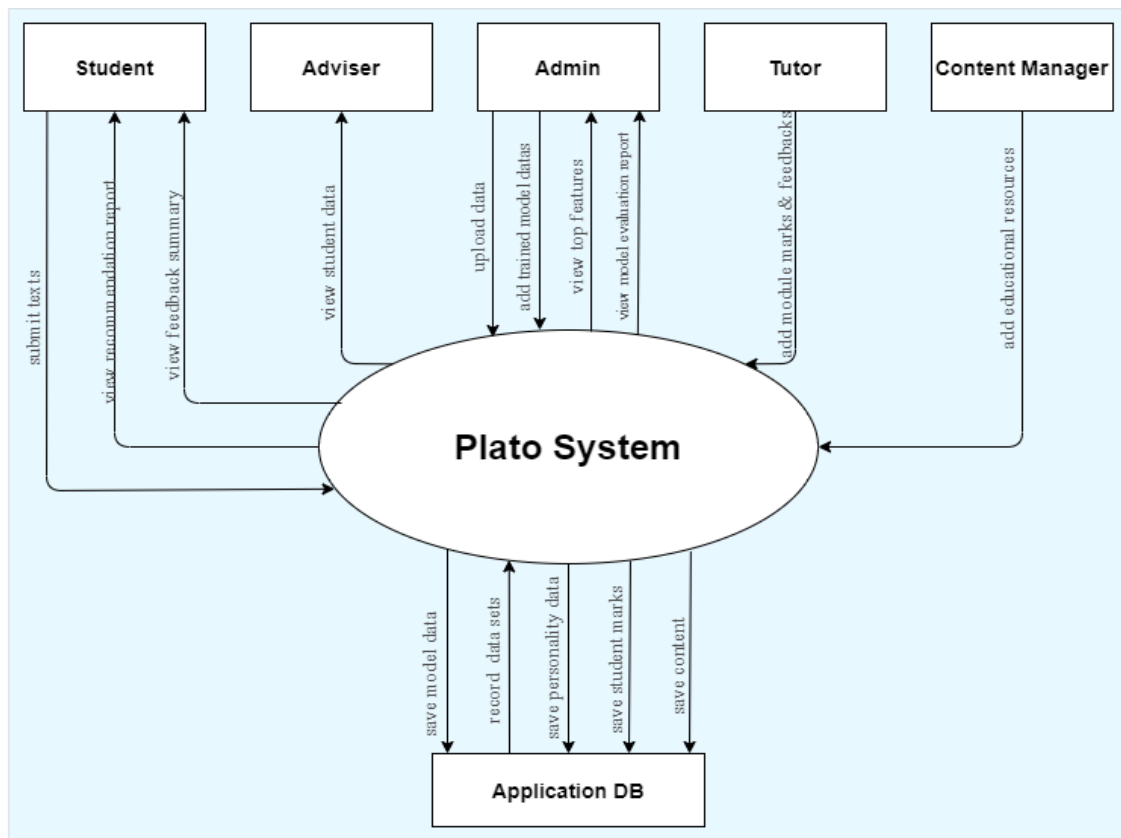
Figure 5. 2: Context Diagram

The Context Diagram shows the context, interaction and boundaries that the system has with external entities, other processes, external data stores. This can be mainly used to identify the flow of the data between internal and external entities.

5.6 Use Case Diagram
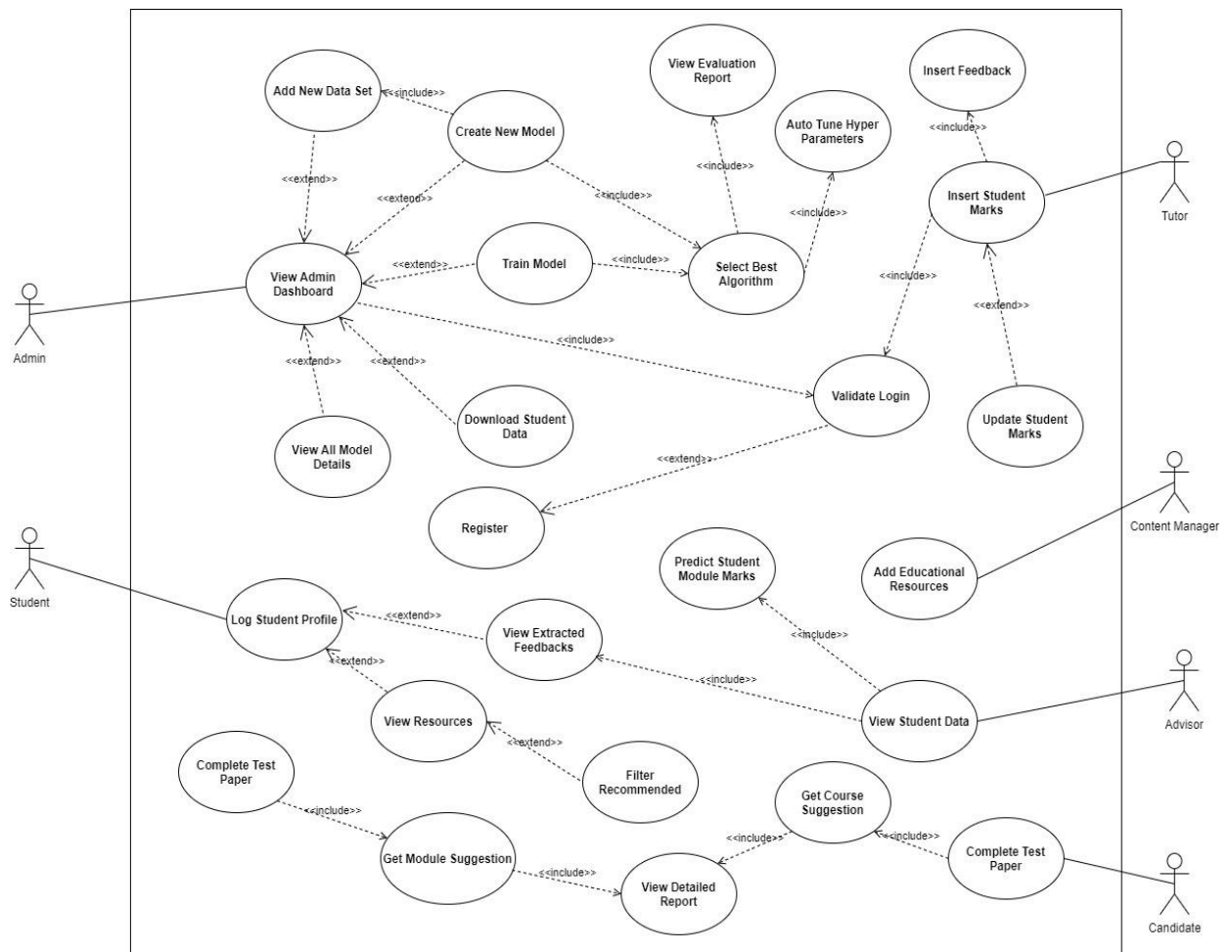
Figure 5. 3: Use Case Diagram

Use Case Id	UC01
Use Case Name	Get path suggestion
Priority	High
Actor(s)	System
Summary	This use case represents the process of generating a path recommendation to the student.

Preconditions	The student should complete the psychometric test.
Extended Use cases	None
Included use cases	View comparison report
Triggered event	Submit a completed psychometric test
Main Flow	Submit completed test Send data to ML server Identify requested prediction type Predict the category or score Generate recommendation report
Alternative flow	None
Exceptional flow	Within the processing period Network connection fails • The use case ends with failure. Display error message
Post condition	Student is able to save the comparison report

Table 5. 6: Use case description 1

Use Case Id	UC02
Use Case Name	Create new model
Priority	High
Actor	Admin, System
Summary	This use case represents the process of training a new model for a module/course by system admin.
Preconditions	User should be logged to the system as an Admin Required Dataset for the model should be uploaded

Extended Use cases	None
Included use cases	Upload New Dataset Select Best Algorithm
Triggered event	Admin Selects create new model for module/course option
Main Flow	Select Create new model
Alternative flow	None
Exceptional flow	Network connection fails Dataset upload fails Model training Interrupts
Post condition	New model created and saves in ML file server.

Table 5. 7: Use case description 2

5.7 Requirement Specification

Identified requirements are prioritised according to three levels as High, Medium and Low in order to implement within resource and time constraints.

Priority Level Identifier	Description
High	Important Core feature of the System
Medium	Not a core feature. But important for the system.
Low	Those can be recognised as 'good if they have' features.

Table 5. 8: Requirement Specification

5.7.1 Functional Requirements

Functional requirements of the Plato are listed below according to the prioritisation among stakeholders.

ID	Description	Priority
FR1	System admin should be able to register students, tutors, advisors, content managers to the system and assign them user roles with different authorities.	High
FR2	Registered users should be able to login the system and they should be redirected to their authorized entry page (ex: profile page or dashboard)	High
FR3	This system should be able to recommend a proper course to pre-registered students by analyzing psychometric tests and classifying them into predefined paths.	High
FR4	This system should be capable of recommending a module to a student by analyzing psychometric tests and predicting their success factor.	High
FR5	The student should receive a detailed report with the recommendation to justify it. All the reasons algorithms consider to make that decision will be included.	High
FR6	The student should be able to request and view the recommended feedback which was given to previous students with similar psychometrics for a particular module.	High
FR7	The student should be able to view and filter recommended educational resources based on their psychometrics.	High
FR8	The system admin should be able to handle data sets by downloading required data sets and uploading appropriate data sets for model training	High
FR9	The system admin should be able to train prediction models with new data and deploy these models by using the available system pipelines	High

FR10	The system admin should be able to view the performance, features of the trained model and other previous model details(confusion matrix , MSE)	High
FR11	The system admin should be able change the prediction, recommendation model type for both internal and public tests as required.	High
FR12	The advisor should be able to view the predicted score for available modules and recommended feedback for a particular student for a particular module.	High
FR13	The tutor should be able to add student marks for particular modules with feedback about the obtained results.	Medium
FR14	The content managers should be able to add suitable educational resources (learning materials) to the system.	Medium

Table 5. 9: Functional Requirements

5.7.2 Non-Functional Requirements

Since the Plato system is designed for the HE institutes it is important to have good actuary, usability and security in the system. Therefore it is crucial to focus on non-functional requirements.

NF1. Accuracy: Accuracy is the most important non-functional requirement in this automated RS. As the end result of the analysed inputs, the recommended course/module should be highly accurate. This should be considered in algorithm selection and using other ML modules in the implementations.

NF2. Performance: The system should rapidly respond based on users' actions under the available workload and system should perform well in all consumer grade major browsers.

NF3. Security: The system should be protected from unauthorized accesses. System should have a proper authentication and authorization mechanism with different levels of access.

NF4. Usability: Plato system should ensure a higher level of usability with students and system administrators by making a usable prototype. User interfaces need to be simple and attractive and responsive to use.

NF5. Maintainability: Plato should be able to maintain by adding or removing required features in an easier way without a major effort of developers.

NF6. Extendibility: This is one of the most important requirements in the Plato system. Since the curriculums and academic pathways are increasing rapidly, the system should be capable of extending its service without redesigning or re developing the system.

NF7. Scalability: After hosting on a server, this system should be scalable enough to assess with the highest workloads and meet the performance requirements since the Plato system architecture is suggesting federated learning mechanism as a future enhancement system should be scalable.

Therefore, this requirement will be considered from early stages of development.

5.8 Chapter Summary

Software Requirements Specification focuses on requirement gathering for the Plato. Started analysing stakeholder engagement with the system. Used Onion model to visualise the level of interaction with the solution of the project. Observing existing systems, literature review, interviews, brainstorming sessions and distributing questionnaires were utilised as the requirement elicitation processes for the system. Findings were summarised to identify the most relevant requirements for the system. Furthermore, stakeholder interactions with the system were defined in use case diagrams, context diagrams analysed actions under the use case description section. The final section explored functional and non-functional requirements based on the priority level which are relevant to the implementation of the system.

CHAPTER 6: TESTING

6.1 Chapter Overview

This chapter presents the various tests which were conducted on the implemented prototype to determine how well it has accomplished its functional and non-functional necessities. Firstly the goals and the objectives testing will be discussed. Then the testing criteria of the Plato will be elaborated. Thereafter Functional requirements testing will be conducted according to the black-box testing mechanism, and finally, the non-functional requirement will be tested.

6.2 Goals and Objectives of Testing

The primary purpose of software testing is to ensure whether the implemented system is working according to the requirements of the software requirement specification. Further, the testing process makes sure the delivery of a stable end product for its target users. Furthermore, the testing process assures the development of the system was followed by the standard practices as it is crucial for maintaining, scaling and extending the system.

6.3 Testing Criteria

The testing process of the Plato will be mainly focused on the following criteria.

- **Test the functionality of data handling, algorithm selection, Predicting marks and comparing courses and modules etc.**

This section will be focused to test whether the core features that require to achieve the research objectives are functioning correctly and providing expected outputs

- **Test accuracy of the models, system performance, security, functional quality etc.**

Since Plato application is designed for non-technical people, the system quality and the architecture of the solution should be at a standard level. Therefore accuracy of the algorithms used to predict the performance, performance, security, usability, maintainability, will be tested, and results will be documented.

The testing criteria mentioned above will be achieved through the following objectives.

1. Validate if the prototype encounters the Functional requirements of the project
2. Validate if the prototype encounters the Non-functional requirement of the project
3. Validate whether the written code follows the best practices and design patterns.

6.4 Functional Testing

All the functional requirements of the system were tested by executing predefined test cases.

ID	Functional Requirement	Steps and Inputs	Expected outcome	Status
TC1	**#FR1** System admin should be able to register users	Add user name, password, email select user role, click Register	A new user should create successfully, and a success message should display.	Pass
TC2	**#FR2** Registered users should be able to log in	Go to the login page, Enter a valid username, password, click login	User should able redirect to particular user type's entry page successfully	Pass
TC3	**#FR3** The system should be able to recommend a course to pre-registered students	Go to the test page without login, complete all test questions, and click View Recommended Course.	Popup should appear with all psychometric test results and course recommendation with the comparison.	Pass
TC4	**#FR4** System should be capable of recommending a module to a student who is already started a course	Log in as a student, open and complete psychometric test, click view recommendations	Popup should appear with psychometric test results, module comparison report and suggestion.	Pass
TC5	**#FR5** Student should receive a detailed report with the recommendation	Complete a psychometric test (internal/ public), click view recommendations	Recommended results should be followed by considered features for the recommendations.	Pass

TC6	**#FR6** Students should be able to view the recommended feedback	Log in as a student, select an upcoming module, click recommend feedback	Summarized feedback should appease for the selected upcoming module.	Pass
TC7	**#FR7** Students should be able to view and filter recommended educational resources	Log in as a student, select a required skill and click view resources, click Filter	Educational resources should be viewed for requested skill and should be filtered for the student.	Pass
TC8	**#FR8** System admins should be able to manage data sets	Log in as an admin, open manage datasets, 1. Select a module and click download. 2. Select a model type and click upload	1. Data set with student psychometrics and module marks for the selected module should download. 2. Data set should upload to the necessary location.	Pass
TC9	**#FR9** System admins should be able to train prediction models	Log in as an admin, open train, create model, select model type, set optimisation type, click train/create	1. Waiting message should appear. 2. After the training completion model hyperparameters/ MSE/ confusion matrix appears.	Pass
TC10	**#FR10** System admins should be able to view the performance, features of the trained model	Log in as an admin, 1.open data analysis, 2.select a model and click view features	1. Previous training data should appear in a grid. 2. Top features of active models should be viewed.	Pass
TC11	**#FR11** System admins should be able to change the internal/public test	Log in as an admin, open change model, go to required test type, select	A success message should appear. Inserted messages should appear at the selected test.	Pass

	result analysing model	model, add a message, and click change.	(Internal test or public test)	
TC12	**#FR12** Advisors should be able to view the student details and the predictions for a module	Login as an advisor, enter student id, select the requested module (available), click view	Student's Psychometric data, predictions and recommendations for the module should be viewed.	Pass
TC13	**#FR13** Tutors should be able to add student marks and feedbacks	Log in as a tutor, Enter required student data, marks and feedbacks	After click Save, a success message should appear.	Pass
TC14	**#FR14** Content managers should be able to add suitable resources	Log in as a content manager, Enter the required resource data	After click Save, a success message should appear.	Pass

Table 6. 1: Test results of functional requirement testing

6.5 System Integration Testing

The Plato system contains various modules to perform various tasks. It is important to test each module separately as well as a complete solution to make sure it is stable as a product. Conducting Integration testing ensures the interface requirements of each module is satisfied. Black box testing approach was followed and each module was tested several times.

Module	Input	Expected output	Actual output	Status
Psychometric questionnaire	Answer list	Report with Psychometric data	As expected	Pass
User registration	User details	Valid account creation	As expected	Pass
Auth Management	Login credentials	Successful redirection to the entry page	As expected	Pass

Course recommendation	Psychometric data	Course comparison report	As expected	Pass
Marks prediction	Student id, module id	Predicted marks	As expected	Pass
Feedback recommendation	Student id, module id	Extracted, summarized feedback	As expected	Pass
Resource recommendation	Student id, skill name	List of filtered educational resources	As expected	Pass
Create a Data set	Module id	CSV file with data set.	As expected	Pass
Create/ train ML models	Main model type, model subtype	Save a pickle file with the trained model	As expected	Pass

Table 6. 2: Test results of integration testing

6.6 Non- Functional Testing

6.6.1 Accuracy Testing

Accuracy

There are few ways to evaluate classification, regression models for having a more accurate idea about the model accuracy without overfitting or underfitting the models. One of those is dividing the data set into three parts as train, validate and test sets. Train and test set will be used to train the model, and the test set will be kept separately to train the model.

Cross-validation is also a widely used technique in ML to test the accuracy of the model and make sure that the model is not overfitting. It takes a portion of data set by test train, splits and trains the model and tests it for each of the positions. Then it takes the average of the measured accuracies. Accuracy testing of the Plato follows the K-fold and other cross-validation mechanisms for each testing process.

First, the available data which contains 221 rows, were converted to 3 data sets. These 3 include 50, 100, 221 rows, respectively. Then these three datasets were used to train GA

with GA-based optimisation. Each training cycle was given around 30 minutes for the optimisation process. Obtained results were as follows.

Data Set Size	50	100	221
Accuracy (%)	36	41	63
Best Pipeline Selected by Tpot Auto ML	DecisionTreeClassifier	GradientBoosting Classifier	KNeighborsClassifier

Table 6. 3: Training cycle was given around 30 minutes for the optimisation process.

```
init(True) #Plato-Main Course Classification Model #GA Opitmization - True
Generation 1 - Current best internal CV score: 0.6268548387096774
Generation 2 - Current best internal CV score: 0.6268548387096774
Generation 3 - Current best internal CV score: 0.6268548387096774
Generation 4 - Current best internal CV score: 0.62747311827957
Generation 5 - Current best internal CV score: 0.62747311827957
Generation 6 - Current best internal CV score: 0.62747311827957
Generation 7 - Current best internal CV score: 0.6276747311827957
Generation 8 - Current best internal CV score: 0.6331048387096774
Generation 9 - Current best internal CV score: 0.6466263440860215
Generation 10 - Current best internal CV score: 0.6466263440860215
Generation 11 - Current best internal CV score: 0.6466263440860215
Generation 12 - Current best internal CV score: 0.6532795698924732
Generation 13 - Current best internal CV score: 0.6532795698924732
Generation 14 - Current best internal CV score: 0.6532795698924732
Generation 15 - Current best internal CV score: 0.6532795698924732
Generation 16 - Current best internal CV score: 0.6532795698924732
```

Figure 6. 1: Figure: GA based optimisation process

Accuracy, Precision, Recall and F1 Score of Classification Models

Accuracy (A) = TP + TN / TP+TN+FP+FN

Evaluation of the classification model should ensure the precision, recall, F1 Score of the model are maintained at an acceptable level. Especially at the scenarios of the data set is an imbalance, these indicators should validate. Furthermore, whether the model needs high precision or recall depends on the purpose of the model.

Precision (P) = TP/ TP + FP   ,   Recall (R) = TP / TP+FN   F1 Score = 2 * (P*R) / P + R

Confusion matrix is providing the values of these basic variables as a model. Classification reports (from sklearn. Metrics) can be identified as a full detailed report about a classification model.

```
: init(True) #Plato-Main Course Classification Model #Loaded From model_Cl.pkl
C:\Users\IsuruP\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\preprocessing\_function_transformer.py:97: FutureWa
rning: The default validate=True will be replaced by validate=False in 0.22.
  "validate=False in 0.22.", FutureWarning)
C:\Users\IsuruP\AppData\Local\Continuum\anaconda3\lib\site-packages\sklearn\preprocessing\_function_transformer.py:97: FutureWa
rning: The default validate=True will be replaced by validate=False in 0.22.
  "validate=False in 0.22.", FutureWarning)
```

```
              precision    recall  f1-score   support

          SE       0.33      0.36      0.34        14
         BIS       0.67      0.69      0.68        26
          NA       0.30      0.25      0.27        12

    accuracy                           0.63        52
   macro avg       0.53      0.53      0.53        52
weighted avg       0.67      0.63      0.64        52
```

Figure 6. 2: Figure: multiclass classification report for main classifier model of Plato

Evaluate Regression Models

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

Root mean squared error values were used to validate all the regression models in the system. For all main regression models root mean squared error was low.

Evaluate Unsupervised Algorithms

KNN: As its output unsupervised KNN is a K number of nearest points and the distance to these points. It doesn't provide any other label. Since the distance to the selected number of nearest points is visible it provides a clear view about the performance of the Algorithm.

Clustering

There are two types of main validation processes to validate quality of the clustering process.

- **Evaluate cluster analysis**: Use unsupervised mechanism to validate the process such as cohesion/separation of the cluster. This can be measured using silhouette coefficient)

- **External evaluation** : Use ground truths, Domain knowledge to validate the clusters

Since there are no ground truths involved with the student clustering process in Plato, an unsupervised method was used to validate the cluster. Created clusters achieved a 0.26 silhouette score which means clusters have clear separations. Here 3 clusters were identified by DBScan.

6.6.2 Performance Testing

Page loading time was tested in google chrome browser which is the world most popular browser, using performance test dev tools in chrome browser.
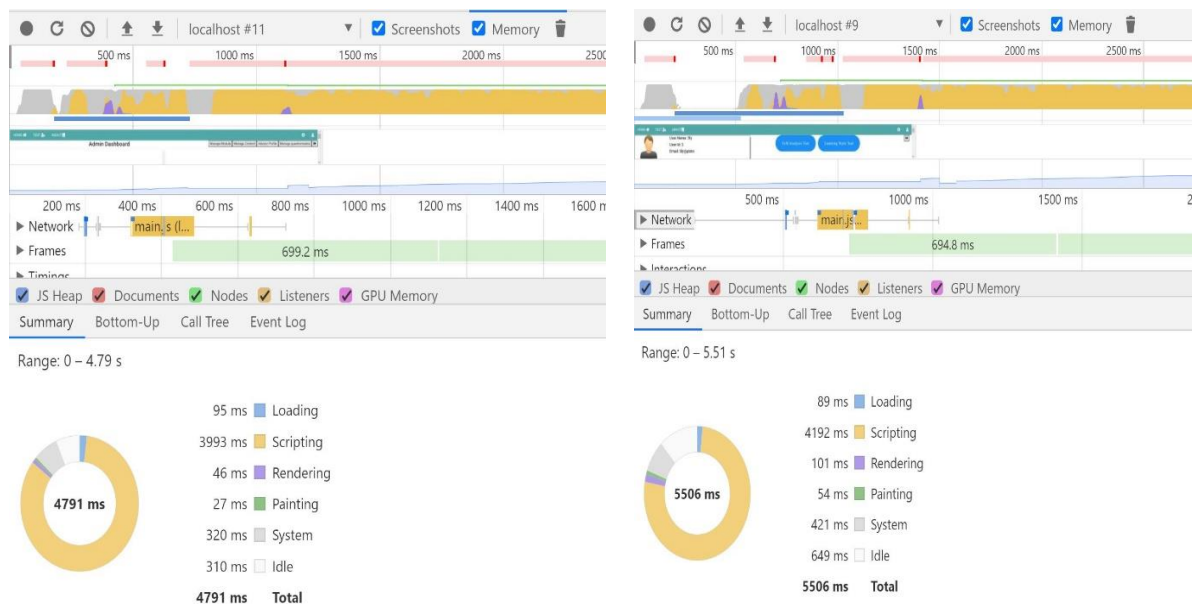
Figure 6. 3: Figure: Loading time from chrome dev-tools

ID	Task	Complete loading time
1	Loading Plato Homepage	5613 milliseconds
2	Loading student profile	5596 milliseconds
3	Loading admin profile	4791 milliseconds
4	Loading public test page	4061 milliseconds

Table 6. 4: Loading time

Load and stress testing

System Performance when handling was mainly tested for the application server.

Select performance test tools

There are several tools to test performance of a web application. Among them apache bench and apache jmeter can be identified as widely used and well recognized tools. Apache JMeter was used to performance testing in Plato public pages
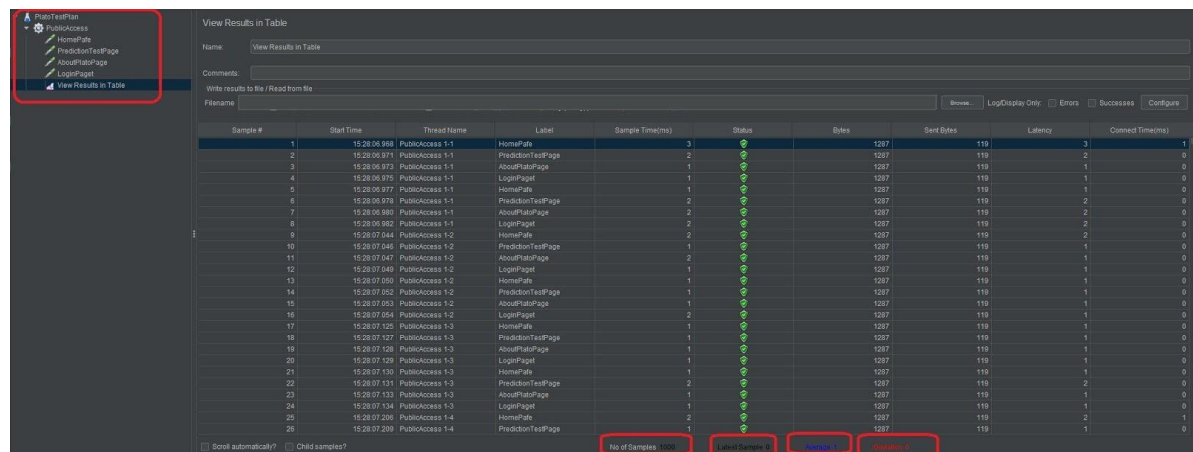
Figure 6. 4: Figure: performance testing in Plato public pages using JMeter

Above graph was generated for 1000 user requests with 10 concurrent rates. System was able to handle it without it failing.

Compatibility of the web application

Proper web based solutions should be compatible with all major consumer garage web browsers such as google chrome, Mozilla Firefox, Microsoft edge etc. This was tested by loading the Plato system in different web browsers and performing user transactions.

Memory Profiling for Application server (Method List)

Below graph shows the memory profiling of view/predict student data for the advisor process which is a comparatively larger process. None of the steps cause to slow the process.

Method	Samples ▼	Own Samples
com.plato.demo.PlatoApplication.main	23	0
jdk.internal.reflect.NativeMethodAccessorImpl.invoke0	22	0
java.lang.reflect.Method.invoke	22	0
org.springframework.boot.SpringApplication.run	22	0
jdk.internal.reflect.NativeMethodAccessorImpl.invoke	22	0
jdk.internal.reflect.DelegatingMethodAccessorImpl.invoke	22	0
org.springframework.boot.SpringApplication.run	22	0
org.springframework.boot.SpringApplication.run	21	0
org.springframework.context.support.AbstractApplicationContext.refresh	21	0
org.springframework.boot.web.servlet.context.ServletWebServerApplicationContext.refresh	21	0
org.springframework.boot.SpringApplication.refreshContext	21	0
org.springframework.boot.devtools.restart.RestartLauncher.run	21	0
org.springframework.boot.SpringApplication.refresh	21	0
org.springframework.beans.factory.support.AbstractBeanFactory$$Lambda$340.2129695335.getObject	12	0
org.springframework.beans.factory.support.DefaultSingletonBeanRegistry.getSingleton	12	0
org.springframework.beans.factory.support.AbstractAutowireCapableBeanFactory.createBean	12	0
org.springframework.beans.factory.support.AbstractBeanFactory.lambda$doGetBean$0	12	0
org.springframework.beans.factory.support.AbstractAutowireCapableBeanFactory.doCreateBean	12	0
org.springframework.beans.factory.support.AbstractAutowireCapableBeanFactory.createBeanInstance	12	0
org.springframework.beans.factory.support.AbstractBeanFactory.doGetBean	12	0
org.springframework.beans.factory.support.AbstractBeanFactory.getBean	10	0
org.springframework.beans.factory.config.DependencyDescriptor.resolveCandidate	8	0
org.springframework.beans.factory.support.DefaultListableBeanFactory.resolveDependency	8	1
org.springframework.beans.factory.support.AbstractBeanFactory.getBean	8	0
org.springframework.beans.factory.support.DefaultListableBeanFactory.doResolveDependency	8	0
org.springframework.boot.web.servlet.context.ServletWebServerApplicationContext.onRefresh	7	0

Figure 6. 5: Memory profiling (view student details to advisor process)

6.6.3 Security

Security is a one of the most important non-functional requirements in the Plato system. Since the system is handling sensitive student data such as module marks this is an essential. The Plato system is protected with a proper authentication and authorization mechanism which was built on spring-security initializer. This ensures security by blocking the access for the users who don't have the access for particular content or who don't have the authority to make specific requests. This can be tested by sending requests without properly login to the system or by trying to access unauthorized content.

This was tested by sending several user requests using the *Postman* tool.

```
1 ▾ {
2       "timestamp": "2020-07-22T18:05:55.697+0000",
3       "status": 403,
4       "error": "Forbidden",
5       "message": "Access Denied",
6       "path": "/register-user"
7 }
```

Figure 6. 6: Security testing

6.6.4 Scalability

This was mainly achieved through the system architecture and by following best practices in the implementation phase.

6.7 Challenges and Limitations

Limitation of the data set was the biggest challenge that phase when testing classification and regression model accuracies. Since the data set contains less number of rows it was hard to test the accuracy for some the classifications and regression models which are using in Plato. Therefore main classification and regression models were used to have an idea about the accuracy. Nevertheless since the Plato admin dashboard provides the facility to view evaluation reports for each model the accuracy of the system, when sufficient data set is available for each use case the accuracy can be maintained in a satisfactory level.

6.8 Chapter Summary

This chapter described the several tests that were performed on the Plato and the obtained results. Functional requirements were tested using black box testing. Further, nonfunctional requirements were done with the help of reliable tools. Accuracy tests were only performed on the main classification models due to the limitation of the dataset. Furthermore, the quality of the developed code was validated.

CHAPTER 7: EVALUATION

7.1 Chapter Overview

The previous chapter explained the testing phase on the implemented prototype of the Plato system. This chapter discusses how the implementation was evaluated by the third parties. Domain specialists, Industrial experts and some of the target audience were selected as evaluators based on various aspects such as development, importance, the usability of the project to provide useful feedback. Selected evaluation criteria, the author's self-evaluation about the completion of the research project are also presented in this chapter.

7.2 Evaluation Methodologies and Approach

Since the Plato is a web-based solution with ML components, which is designed to provide solutions to the HE domain by using educational psychological factors, it covers the few separate spheres. Therefore it is important to find the evaluators from each related field. Evaluators are mainly divided into 2 separate groups as technical (industrial experts, software engineers) and non-technical (domain specialists, end-users). Below documents were shared with every evaluator.

- Introduction and project demonstration of Plato: https://youtu.be/4Ru833o8inM
- Overview explanation website: http://platoguide.s3-website-us-west-2.amazonaws.com

In addition to this, an online demonstration session was conducted with each Industrial expert, software engineer to thoroughly discuss the system architecture, technology stack and implementation process.

7.3 Evaluation Criteria

Concept of the Project	To assemble comments on the prototype to understand the reception from the domain and industrial experts
Scope and the depth of the project	To assure if the scope of the project is broad enough
Design, architecture and implementation	To assure that the project is able to achieve its architectural design goals.

Solution and implemented prototype	To make sure if the prototype is good enough to represent the proposed solution
Identified limitations and possible Future Enhancements	To obtain insights and comments of the experts about the aspects that should consider in the next phases
General feedback	To gather any other general commentary or an idea about the research project system or prototype

Table 7. 1: Evaluation Theme

7.4 Selected Evaluators

Education and Psychology Domain Experts

Name	Description
UOW lecturer	A Principal Lecturer/ Course leader (UOW, UK)
Inoka Samarasekara	Senior Lecturer, University of Moratuwa, Sri Lanka,PhD Researcher at University of New South Wales, Australia.
Dr. Kanthi Hettigoda	Clinical Psychologist, PhD in psychology, Done several researches regarding educational psychology
KAV Abeygunawardena	Phd Research Scholar, Done several Researches in Education domain including educational choices ,decision making

Table 7. 2: Evaluator Selection

Industrial Experts and Software engineers

Name	Description
Uditha Nagahawatta	Chief Architect - Global Market Technologies.
Chaminda Hettigoda	Software Architect (Global market technologies)

	Have more than 15 years' experience in developing transaction processing systems using Java and JavaEE.
Parakrama Rathnayake	Lead Product Owner (Pickme foods) 15 years of experience in product development (BA, PO) stream.
Amila Manathunga	Software Architect (Global market technologies)
Lahiru Jayakodi	Software Engineer/ ML researcher (GMT)

Table 9. 3: Evaluator Selection

Students represent possible end-users

Purushoth Anantharasa	BEng Software Engineering Undergraduate (IIT)
Randul Malinhara	Foundation Course Student at IIT
Pasan Gunawardena	School Leaver (D.S. Senanayake College, Colombo)

Table 7. 4: Evaluator Selection

7.5 Evaluation Results

7.5.1 Overall Concept

UOW lecturer	It's very interesting and very useful to support student choice of modules, courses and direction of their future career
Inoka Samarasekara	Overall concept is interesting and can be considered as a current requirement across the HE sector.
Dr. Kanthi Hettigoda	Very good and pragmatic
KAV Abeygunawardena	Good. Because it improves the students' decision-making process.

Uditha Nagahawatta	This is a really good initiative to improve the efficiency at institutional level by encouraging students to follow courses where they are good at. Also this will let students to understand where their skills really are. Concept of the product is well thought through and implemented.
Chaminda Hettigoda	Excellent
Parakrama Rathnayake	Can see a clear use-case. i.e. a clear pain point in the market will be addressed through this solution.
Amila Manathunga	It is an excellent choice. To be honest I get many phone calls from different sorts of young individuals to get career guidance. This proves that selecting the best and matching career is a real problem in the society. As far as I know, there are very limited tools available to guide these newbies. So this tool is a very demanding requirement in the modern era. I should thank Sapumal for selecting such a demanding and challenging project.
Lahiru Jayakodi	Good. Valid hypothesis.
Purushoth Anantharasa	The project is really helpful for students who are struggling in their degrees without proper guidance.
Randul Malinhara	As this project mentioned, I'm definitely sure this will help the student like us to show the better path to choose in Higher Edu.
Pasan Gunawardena	This is a project that is essential for the present and will continue to be useful in the future.

Evaluation Summary: As the feedback clearly shows all the evaluators appreciate the overall concepts since it is trying to address a demanding requirement in current society.

Table 7. 5: Summary of evaluation feedback on the Overall Concept

7.6.2 Scope of the Project

UOW lecturer	I thought the psych-metrics analysis together with their academic performance to find the future fit is a good approach.
Inoka Samarasekara	The scope and the depth of the project is adequate and satisfactory.
Dr. Kanthi Hettigoda	Adequate for this level
KAV Abeygunawardena	Implementation prototype seems to be covering the scope and depth is highly recommended for BEng.
Uditha Nagahawatta	Scope and the depth of the project is more than enough for a single person. The student has researched and implemented the project to usable end product level.
Chaminda Hettigoda	He has covered the large scope and depth required for a final year project.
Parakrama Rathnayake	Covering the whole product scope is impressive.
Amila Manathunga	Scope and the depth is at an exceptional level.
Lahiru Jayakodi	Scope and depth can be narrower than this
Evaluation Summary: All the evaluators were totally satisfied with scope and the depth of the project.	

Table 7. 6: Summary of evaluation feedback on the Scope of the Project

7.6.3 Design, Architecture and Implementation

This criteria was evaluated by the technical evaluators

Uditha Nagahawatta	Very nicely done by using industry standard tools and concepts.
Chaminda Hettigoda	Design the system using software engineering concept and latest technologies
Parakrama Rathnayake	As per the available information design is well above average.
Amila Manathunga	As an academic project, the design and architecture is at an exceptional level. But when considering industry standards, there are few areas to improve such as efficient asynchronous data loading, more improved UX. Again, I should really appreciate the level of details addressed in the project considering the level of an academic project.
Lahiru Jayakodi	Really appreciate you have a working prototype with a capability to enhance. High-level project design and architecture pointed to the correct direction.
Evaluation Summary: This criteria was thoroughly evaluated by all the industrial experts and software engineers. All of them were highly satisfied about the implementation of the project, decisions made, and the tech stack used. Especially the design and the architecture of the solution were identified and appreciated as above average.	

Table 7. 7: Summary of system Design, Architecture, and Implementation feedback

7.6.4 Implemented Solution features and Prototype

UOW lecturer	It has great potential in the self-help arena and possibly a tool for career advice services.

Inoka Samarasekara	It is vital to validate the solutions with existing data. For example, overlaps and exceptions can be found in personality types. Providing the other possible options in terms of the suitable pathways/modules might be a good solution.
Dr. Kanthi Hettigoda	This meets a timely requirement of modern education
KAV Abeygunawardena	As understood, the solution depends on the historical data available in the system and within the implementation process expert advisory recommended with the output provided by the solution Features implemented at present will be a good starting point and more features can be considered with the feedback after implementation.
Uditha Nagahawatta	The student has implemented it beyond prototype level. What requires more is to feed considerable amount of actual data into the system to improve the accuracy.
Chaminda Hettigoda	As a prototype system is working properly and covers the scope of the project. He has developed it so that it can easily configure and extend.
Parakrama Rathnayake	Well thought through and planned. It's done in a way that it can easily be extended for any future requirements.
Amila Manathunga	Looks very matured and a considerable thought process has gone through the solution, performance and the PT.
Lahiru Jayakodi	You have implemented more than enough features. The way you addressed the problems in ML is most reasonable. Need to focus more on model scores

Purushoth Anantharasa	The Prototype has a good and understanding User Interface for a user to quickly capture the functions and use it effectively.
Randul Malinhara	The students or any other person who wished to select their path wisely, I highly recommend this solution because this will give a confidence about the path which was decided by he/she.
Pasan Gunawardena	Designed in a way that is useful and very easy to guide for students preparing for HE at the moment.

Table 7. 8: Feedback summary for Implemented Solution features and Prototype

7.6.5 Limitations and Future Enhancements

UOW lecturer	I feel the data protection and security aspects should be emphasised. There is a lot of personal and private data being provided in using the system.
Inoka Samarasekara	It would be important to display the prerequisites to follow a particular pathway/module
Dr. Kanthi Hettigoda	There are other psychological or personality characteristics which may contribute to their education decision. However he has covered enough
KAV Abeygunawardena	limited historical at the beginning/ until sufficient data collected to guide students and stakeholders, the prediction is limited to the selected models and weights
Uditha Nagahawatta	Some areas of the web UI needs to be enhanced to bring it to a sellable level. The only limitation I can see is not having access to large amount of actual data to improve the accuracy of the output.

Chaminda Hettigoda	Increase the data set used to get the decision
Parakrama Rathnayake	Better to plan how this solution could be used to help students to pick an Institute or a subject stream.
Amila Manathunga	I only see some improvements in the UX aspect where making asynchronous data loading more organised. However from my experience, I should rate this project at a very high level considering the quality and outputs in an academic project context.
Lahiru Jayakodi	Regarding the ML solution, you need to have a correct number of data. Instead of event driven model training, you might have to do it as a scheduled batch jobs. Regarding your implementation, you have room to improve UX in front-end designs. Think from a user point of view.
P. Anantharasa	Not Identified
Randul Malinhara	Nothing. because in every aspect which can guide a student to choose his/she path, have it completed by this system.
Pasan Gunawardena	Nothing at the moment
Evaluation summary: Evaluators highlighted some UI. UX related enhancements. One evaluator noted that security aspects related to the privacy data.	

Table 7. 9: Feedback for Limitations and Future Enhancements

7.6.6 General Feedback about the research project and prototype

UOW lecturer	Very interesting and useful tool

Inoka Samarasekara	A good project with rooms for improvement. Wish you all the very best and keep it up.
Dr. Kanthi Hettigoda	Good effort and unique features
KAV Abeygunawardena	Very good starting point. Wish you good Luck !
Uditha Nagahawatta	If this can be used at institutional level, will help students a lot to efficiently select their courses. I encouraged the student to take this product beyond academic to a production level, at least to one institute.
Chaminda Hettigoda	One of the excellent final year project I have seen
Parakrama Rathnayake	Overall it's a nice concept, thus I see a potential market value. Product covers almost all the angles the solution has. So it's without a doubt, qualified for a good MVP.
Amila Manathunga	Excellent work and I'm sure this concept will be developed into a more matured commercial product which would address a burning problem in the modern day society.
Lahiru Jayakodi	You have touched a lot of areas and approaches in ML. Really appreciate seeing such voluntary motivation and effort from a student. Along with many students who try to get projects done with minimum involvement, I see you as an exceptional student.
Purushoth Anantharasa	Plato is an excellent idea for students who are struggling without proper guidance in non relevant degree programs.

Randul Malankara	As a student in HE, We need this kind of a Guidance to choose the best path to win the future.
Pasan Gunawardena	I think this project can increase the productivity of the field of education by providing it to students who wish to pursue HE as soon as possible.

Table 7. 10: General Feedback

Based on the overall feedback, evaluators are satisfied with the project goals an encouraged for the practical implementation in universities. Further expressed wishes for future continuations.

7.6 Evaluation on Completion of Functional Requirements

ID	Description of functional requirement	Priority	Status
FR1	The system admin should be able to register users	Critical	Completed
FR2	Registered users should be able to log in	Critical	Completed
FR3	The system should be able to recommend a course to pre-registered students	Critical	Completed
FR4	The system should be capable of recommending a module to a student who is already started a course	Critical	Completed
FR5	The student should receive a detailed report with the recommendation	Critical	Completed
FR6	The student should be able to request and view the recommended feedback	Critical	Completed
FR7	The student should be able to view and filter recommended educational resources	Critical	Completed
FR8	The system admins should be able to manage data sets	Critical	Completed
FR9	The system admins should be able to create and train prediction models using the system pipelines	Critical	Completed

FR10	The system admins should be able to view the performance, features of the trained model	Critical	Completed
FR11	The system admins should be able to change the internal/public test result analysing model	Critical	Completed
FR12	The advisor should be able to view the student details and the predictions for a module	Critical	Completed
FR13	The tutors should be able to add student marks, feedbacks	Important	Completed
FR14	The system content managers should be able to add suitable educational resources	Important	Completed

Table 7. 11: Reflection of Completion of Functional Requirements

7.7 Evaluation of completion Non-Functional Requirements

Requirement and Evaluation	Priority	Status
NF1. Accuracy Most optimum model and hyperparameter selection by pipelines in each training cycle. Validate the confusion matrix/ MSE after every training sequence.	Critical	Achieved
NF2. Performance Provide requested solutions using minimum steps possible. Ability handles considerable concurrent user requests.	Critical	Achieved
NF3. Security Different user levels. Proper Authentication and authorization mechanism to validate every request to the application server.	Critical	Achieved
NF4. Usability Good UI design UX concepts in system interface, Proper GUI to perform data science tasks. Well tested in major browsers.	Important	Achieved
NF5. Maintainability	Important	Achieved

Feasibility to add, generate, use all the required data using the system itself. Proper code structure.		
NF6. Extendibility The mechanism to add, change prediction engines using a dashboard without changing or restructuring the system.	Important	Achieved
NF7. Scalability Proper system architecture, Use appropriate design patterns in the development phase.	Important	Achieved

Table 7. 12: Reflection of completion Non-Functional Requirements

7.8 Self-Evaluation

Evaluation Criteria	Self-Evaluation
Overall Project Concept	This project concept originated with the intention of providing a viable solution for a major problem in the HE domain. Further, the concept was developed with proper research in the domain. The domain experts and researchers in the HE domain appreciate the concept as a good solution for the particular problem
Scope of the project	According to the selected domain and the research question of this project, the research should be conducted on a considerably wider area. When considered the industrial best practices to complete a functional prototype, it covers the best possible scope and the research depth, which is likely to be achieved in a limited duration.
Architecture and Design of the system	The design and the architecture of the system was focused to provide a great user experience to users and mainly to protect the maintainability, scalability and extendibility of the system.

Features and performance in prototype	The system prototype was developed up to an enterprise-level application with all necessary functional and non-functional requirements by following programming best practices. Therefore after conduction training cycles with a large data set and after a few further improvements and fine-tunes, the prototype can be deployed for the production to test in an actual HE institute.
Limitation and future enhancements	The main limitation of this kind of system is enough data to make a more accurate decision in each scenario. After planting this system in a HE domain it may take some time to gather all required data before stating the prediction and recommendation process. Possible solutions for this and future enhancements will be discussed in the conclusion chapter.

Table 7. 13: Self-Evaluation

7.9 Chapter Summary

This chapter covered the documentation part of the evaluation phase of the project. The findings and evaluations stated in this chapter are crucial to take an overall understanding of the completion and the success of the research project. In the beginning, the evaluation was selected and structured according to the requirements of the project completion. Then the evaluator selection process and prototype demonstration process were documented. Furthermore, the functional and nonfunctional requirement completion were evaluated by the author. Finally, the author evaluated the conducted research as a whole

CHAPTER 8: CONCLUSION

8.1 Chapter Overview

This chapter contains the final remarks and the reflections of this research project. First, it reflects on the aim and objectives of the project and whether they are successfully achieved. Then it discussed the utilization of the knowledge and experience of the author for this project. Next, the knowledge gained by conducting this research is included. Moreover, the problems and challenges faced through the research path are discussed. Finally, the possible future enhancements and the impact on the research community and the educational domain is described.

8.2 Achievement of Project Aim and Objectives

8.2.1 Achievement of Research Objectives

Objective Description	Status
To perform an in-depth literature review in HE, educational psychology domain and machine learning domain to identify a proper method to identify the possible approaches and mechanisms to achieve the research aim	Completed
To analyse the requirements for a proper system which is capable of providing recommendations to make effective decisions through the HE pathway.	Completed
To develop a system which is capable of filling existing gaps and fulfil the requirements of possible end-users in the HE domain.	Completed
To evaluate the developed prototype to make sure the system is capable of meeting the requirements of end-users, domain experts and industry.	Completed

Table 8. 1: Achievement of Research Objectives

8.2.2 Achievement of the Aim

To design, develop and evaluate a system which is capable of helping students by providing recommendations to make more accurate and useful decisions throughout the academic

pathway by considering their psychometrics, and at the same time can meet the industrial requirements as a web-based solution. Above mentioned research aim was successfully achieved by designing and developing the. Plato student guidance system and evaluating it by the domain specialists, industry experts and end-users. Successful completion of the research objectives accomplished the research aim.

8.3 Utilising of Knowledge from Course Content

Software development principles I and II	The author stepped into the programming through these modules. Several basic principles and fundamentals of programming were learned from these modules and that knowledge was helped in many ways for this project.
Object-oriented programing	This module introduced the author to OOP and used them to increase the flexibility and effectiveness of the system.
Database, Information systems	Database architecture design, managing integrity, querying efficiently are required aspects of the project. Fundamental of these were learned from these modules.
Web Technology	How to design and develop a basic web page and style those how to perform validation, basic ways to communicate with a backend were introduced by this module. Since Plato application should provide a full-featured UI web development knowledge which begins from this module was quite useful.
Client server architecture	This module guided how to design and develop a complete client-server architecture and taught several principles which are important to consider when moving along with client-server communication mechanisms.
Software development group project	The way of initiating and conducting a project from the design to functionality testing was learnt through this module. Additionally, it provides a good experience in the project documentation.

Table 8. 2 : Utilising of Knowledge from Course Content

8.4 Use of Existing Skills/ Experience

Industrial experience as a full-stack web, mobile developer: The author has worked in the software industry for a few years as a full-stack web, mobile developer. At this period author was involved in every phase of the software development life cycle in large scale

international level financial solutions. The experience and skills gained through this were very helpful to plan and execute a project like this. When designing the solution architecture it was planned by using the previous experience of planning scalable and extendable solutions.

Experience as a volunteer counsellor: The author has some experience as a volunteer counsellor for an international non- profit online organization. This was a good experience to understand about the student psychology and existing problems in the HE domain since many students are seeking some help to make their mind up when having no preference to continue their studies in a previously selected pathway. This was one of the key factors to motivate the author to design and develop this kind of a system as a usable product.

Experience as a teacher: Author has some working experience as a volunteer teacher for some school and high school level students in a few subjects. This also helps the author to have an understanding of the education domain.

8.5 Learning Outcomes

Conducting a full-scale research project: This can be identified as the main learning outcome of this project. These kinds of projects need to identify and analyze lots of factors. Further many logical decisions have to be made. This may be an essential lifelong skill.
Technical skills: This research project gave the first experience in data science. It helps the author to obtain many skills such as having an intuition about the behaviour of various ML algorithms can be identified as the most important one. Further, these projects provide the experience from collecting a dataset to deploying and hosting an ML model successfully.
Educational psychology: As a result of conducting in-depth research in this domain and discussing with domain specialists, the author obtained a good understanding of psychology and educational psychology which may be worthy for any career path.
Technical writing/ formal documentation: Completing this thesis provided a great experience to the author in terms of technical documentation.
Life skills: This project assists the author to improve many areas such as time management, critical thinking, problem-solving, analyzing and decision making.

Table 8. 3: Learning Outcomes

8.6 Limitations Challenges and problems faced

Finalize the key features of the system: This research can contain a huge area of the educational domain, psychology and machine learning algorithms. Also, this kind of system can contain lots of important features but with the time limitation and lack of data, it was important to finalize key functionalities to implement. After several discussions with the supervisor, identified the major key functionalities to implement.

Define the Scope: Same as the above key features it was important to finalize a scope of a project like this. The scope should be large enough to show all the key functionalities and should be possible to reach with limited time and limited data. With the help of the supervisor, the scope was decided.

Find the data set: For the project plan, it is a must to gather student personality data, brain dominance data, past performance and the current exam marks of the same student. There are many rules and regulations in universities like they will not provide any student data to a third party. This was a huge challenge to face. As a solution for that customized questionnaire component designed to provide students individually and gather data.

Lack of implementation knowledge: Since this project deals with advanced ML algorithms and concepts initial knowledge of ML concepts were not enough. After identifying this challenge, started to gather more knowledge by referring to online tutorials and books.

Problems faced due to Covid-19: This affected the completion of the project in a few ways. First, the data set gathering process was unable to conduct successfully as planned previously. It also affected the expected evaluators. But mainly the personal impact to the author was the biggest difficulty faced. As a result of some personal reasons, the author came to a point that was unable to continue this research project. As a result of that, the author requested a deferral.

8.7 Future Enhancements

Essential Enhancements

#1	Improve the accuracy with a sufficient data set.
#2	Add more user-friendly and detailed questionnaire to measure psychometric variables
#3	Run complete test cycles on different environments and add all the necessary validation and notifications in each possible scenario.
#4	Improve security by adding encryptions and other important mechanisms
#5	Create a complete system documentation

Table 8. 4: Essential Future Enhancements

Enhancements to improve the solution and provide additional features

#6	Add deep learning models for well tested and established recommendation and prediction tasks. After the system is matured enough this can be considered.
#7	Add Federated learning to the system: Federated learning may be an essential concept in the education domain in future. As author research and identifies in the literature review chapter that may be capable of providing various rapid improvements in educational data mining. It lets Plato like systems the ability to recognize patterns in students' behaviour on a large scale compared to courses at the inter-university level.
#8	Predict student psychological status and state change probabilities using Markov models.
#9	Implement and use components with different data sets such as learning styles to make more usable decisions.
#10	Make the feedback summarizer abstractive other than extractive
#11	Question pool for the psychometric tests to generate various tests

Table 8. 5: Enhancements to improve the solution and provide additional features

8.8 Reflections on Contribution to the problem domain

This research shows a novel and feasible way to create a university-based recommendation system which has the ability to provide recommendations throughout the academic pathway. In addition to creating a RS, this research mainly focused on solving the challenges that such a RS can face when it is used in an actual environment.

8.9 Concluding Remarks

This marks the end of this documentation and the research project. It was a remarkably memorable journey for developing a system as a solution for a real-world problem. Further, this effort and the delivered product were appreciated by several domain specialists and industry experts. It indicates that the research project can be concluded with satisfaction. Finally, even though this research is concluded, the author expects to continue the work on making improvements in the educational domain

REFERENCES

1. Abel, J. and Deitz, R. (2014). Do the Benefits of College Still Outweigh the Costs?. CURRENT ISSUES IN ECONOMICS AND FINANCE, 20 (3), 4,5. Available from https://www.newyorkfed.org/medialibrary/media/research/current_issues/ci20-3.pdf [Accessed 14 February 2020].

2. Agrawal, H. and Mavani, H. (2015). Student Performance Prediction using Machine Learning. International Journal of Engineering Research and Technology, 4 (03). Available from 10.17577/ijertv4is030127 https://www.ijert.org/research/student-performance-prediction-using-machine-learning-IJERTV4IS030127.pdf [Accessed 12 November 2019].

3. Ali, J., Khan, R., Ahmad, N. and Maqsood, I. (2012). Random Forests and Decision Trees. International Journal of Computer Science Issues, 9 (5), 272-278. Available from https://www.semanticscholar.org/paper/Random-Forests-and-Decision-Trees-Ali-Khan/959a8e906ee26b940374b719253c8e188ed78fd3 [Accessed 14 November 2019].

4. Altun, A. (2012). Ontologies for Personalization: A New challenge for Instructional Designers. Procedia - Social and Behavioral Sciences, 64, 691-698. Available from https://www.sciencedirect.com/science/article/pii/S1877042812050586 [Accessed 7 February 2020].

5. Al-Zoubi, S. and Younes, M. (2015). Low Academic Achievement: Causes and Results. Theory and Practice in Language Studies, 5 (11), 2262. Available from 10.17507/tpls.0511.09 https://www.semanticscholar.org/paper/Low-Academic-Achievement%3A-Causes-and-Results-Alzoubi-Younes/44eac76565236250bbce0071c4ca0f5c64299f3c [Accessed 16 October 2019].

6. Asif, R., Merceron, A., Ali, S. and Haider, N. (2017). Analyzing undergraduate students' performance using educational data mining. Computers & Education, 113, 177-194. Available from 10.1016/j.compedu.2017.05.007 https://www.sciencedirect.com/science/article/abs/pii/S0360131517301124 [Accessed 5 October 2019].

7. Balaji, A. and Allen, A. (2018). Benchmarking Automatic Machine Learning Frameworks. Available from https://arxiv.org/abs/1808.06492 [Accessed 25 February 2020].

8. Baradwaj, B. and Pal, S. (2011). Mining Educational Data to Analyze Students" Performance. International Journal of Advanced Computer Science and Applications,, 2 (6), 63-69. Available from https://arxiv.org/ftp/arxiv/papers/1201/1201.3417.pdf [Accessed 25 February 2020].

9. BAKER, R. and YACEF, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1 (1), 3-16. Available from https://eric.ed.gov/?id=EJ1115366 [Accessed 20 November 2019].

10. Bhardwaj, B. and Pal, S. (2011). Data Mining: A prediction for performance improvement using classification. International Journal of Computer Science and Information Security, 9 (4). Available from https://arxiv.org/ftp/arxiv/papers/1201/1201.3418.pdf [Accessed 22 November 2019].

11. Bhila, T. (2018). The Benefits and Generic Procedure of Automating an Academic Student System in Primary and Secondary Schools as an Impetus for Educational Technology. International Journal of Innovative Science and Research Technology, 3 (11). Available from https://ijisrt.com/the-benefits-and-generic-procedure-of-automating-an-academic-student-system-in-primary-and-secondary-schools-as-an-impetus-for-educational-technology [Accessed 2 November 2019].

12. Biswas, S. (2018). Choices Based Credit System(CBCS)— An analytical study. International Journal of Research and Analytical Reviews, 5 (3). Available from http://ijrar.com/upload_issue/ijrar_issue_1469.pdf [Accessed 10 February 2020].

13. Bjorklund-Young, A. (2016). *What Do We Know About Developing Students' Non-cognitive Skills?*. What Do We Know About Developing Students' Non-cognitive Skills?. Available from http://edpolicy.education.jhu.edu/wp-content/uploads/2016/06/NoncognitiveskillsmastheadFINAL.pdf [Accessed 4 March 2020].

14. Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., Kiddon, C., Konecny, J., Mazzocchi, S., McMahan, H., Petrou, D., Ramage, D. and Roselander, J. (2019). Towards Federated Learning at Scale: System Design. arXiv,

1-15. Available from https://arxiv.org/pdf/1902.01046.pdf [Accessed 14 November 2019].

15. Buehring, S. (2018). PRINCE2® processes. Knowledge Train. Available from https://www.knowledgetrain.co.uk/res/ebooks/prince2-processes-ebook.pdf [Accessed 16 March 2020].

16. Ceja, M. (2006). Understanding the Role of Parents and Siblings as Information Sources in the College Choice Process of Chicana Students. Journal of College Student Development, 47 (1), 87-104. Available from 10.1353/csd.2006.0003 https://eric.ed.gov/?id=EJ743914 [Accessed March 2020].

17. Chamorro-Premuzic, T. and Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. Journal of Research in Personality, 37 (4), 319-338. Available from https://www.sciencedirect.com/science/article/abs/pii/S0092656602005780?via%3Dihub [Accessed 8 December 2019].

18. Dutt, A., Ismail, M. and Herawan, T. (2017). A Systematic Review on Educational Data Mining. IEEE Access, 5, 15991-16005. Available from 10.1109/access.2017.2654247 https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7820050 [Accessed 5 February 2020].

19. Eidimtas, A. and Juceviciene, P. (2014). Factors Influencing School-Leavers Decision to Enrol in Higher Education. 5th World Conference on Educational Sciences - WCES 2013, 116, pp.3983-3988. Available from : https://www.sciencedirect.com/science/article/pii/S1877042814008945 [Accessed 25 October 2019].

20. Sustainable Development Goals (2020). Eua.eu. Available from https://eua.eu/issues/24.html [Accessed 9 February 2020].

21. Fallan, L. (2006). Quality Reform: Personality type, preferred learning style and majors in a business school. Quality in Higher Education, 12 (2), 193-206. Available from 10.1080/13538320600916817 https://eric.ed.gov/?id=EJ753640 [Accessed 7 October 2019].

22. Feurer, M. and Hutter, F. (2018). Automatic Machine Learning: Methods, Systems, Challenges. Available from https://www.automl.org/wp-content/uploads/2018/12/automl_book.pdf [Accessed 26 March 2020].

23. Granik, M. and Mesyura, V. (2017). Fake News Detection Using Naive Bayes Classifier. 2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON), 900-903. Available from https://ieeexplore.ieee.org/document/8100379 [Accessed 19 November 2019].

24. Gayathri, K. and Marimuth, A. (2016). HYBRID ONE-PASS CLUSTERING WITH ENHANCED KNN CLASSIFIER COMBINED WITH EM FOR TEXT CATEGORIZATION (H3TC) ALGORITHM. International Journal of Science & Technology, 6 (3), 35-42. Available from http://www.ijst.co.in/papers/vol6issue3/ijst_160704.pdf [Accessed 11 March 2020].

25. Genovese, J. (2005). Hemispheric Cognitive Style: A Comparison of Three Instruments. The Journal of Genetic Psychology, 166 (4), 467-481. Available from 10.3200/gntp.166.4.467-482 https://pubmed.ncbi.nlm.nih.gov/16463609/ [Accessed 4 March 2020].

26. Gray, G., McGuinness, C., Owende, P. and Carthy, A. (2014). A Review of Psychometric Data Analysis and Applications in Modelling of Academic Achievement in Tertiary Education. Journal of Learning Analytics, 1 (1), 75-106. Available from 10.18608/jla.2014.11.5 https://eric.ed.gov/?id=EJ1127032 [Accessed 3 January 2020].

27. Gulati, J., Bhardwaj, P., Suri, B. and Lather, A. (2016). A Study of Relationship between Performance, Temperament and Personality of a Software Programmer. ACM SIGSOFT Software Engineering Notes, 41 (1), 1-5. Available from 10.1145/2853073.2853089 https://dl.acm.org/doi/abs/10.1145/2853073.2853089 [Accessed 8 December 2019].

28. Gupte, A., Joshi, S., Gadgul, P. and Kadam, A. (2014). Comparative Study of Classification Algorithms used in Sentiment Analysis.) International Journal of Computer Science and Information Technologies, 5 (5), 6261-6264. Available from https://pdfs.semanticscholar.org/4667/88e0ba1f608981ca5422ddfb5bfedeef75d0.pdf [Accessed 15 October 2019].

29. Halde, R., Deshpande, A. and Mahajan, A. (2016). Psychology assisted Prediction of Academic Performance using Machine Learning. IEEE International Conference On

Recent Trends In Electronics Information Communication Technology, 431-435.
Available   from   10.1109/RTEICT.2016.7807857
https://ieeexplore.ieee.org/document/7807857 [Accessed 12 November 2019].

30. Hamsa, H., Indiradevi, S. and Kizhakkethottam, J. (2016). Student Academic
    Performance Prediction Model Using Decision Tree and Fuzzy Genetic Algorithm.
    Procedia Technology, 25, 326-332. Available from 10.1016/j.protcy.2016.08.114
    https://www.sciencedirect.com/science/article/pii/S2212017316304613 [Accessed 5
    November 2019].

31. Harrington, C. and Orosz, T. (2018). Why the first-year seminar matters, 1st ed.
    London: Rowman & Littlefield.

32. Hull-Blanks, E., Kurpius, S., Befort, C., Sollenberger, S., Nicpon, M. and Huser, L.
    (2005). Career Goals and Retention-Related Factors Among College Freshmen.
    Journal of Career Development, 32 (1), 16-30. Available from
    https://journals.sagepub.com/doi/10.1177/0894845305277037 [Accessed 20 October
    2019].

33. Hyde, K. (2012). Qualitative Market Research: An International Journal. Qualitative
    Market Research: An International Journal, 3 (2), 82-89. Available from
    https://www.emerald.com/insight/content/doi/10.1108/13522750010322089/full/html
    [Accessed 18 January 2020].

34. Irani, T., Telg, R., Scherler, C. and Harrington, M. (2003). PERSONALITY TYPE
    AND ITS RELATIONSHIP TO DISTANCE EDUCATION STUDENTS' COURSE
    PERCEPTIONS AND PERFORMANCE. The Quarterly Review of Distance
    Education,   4   (1528-3518),   445-453.   Available   from
    https://www.semanticscholar.org/paper/Personality-Type-and-Its-Relationship-to-
    Distance-Irani-Telg/b73d12befcb0da1abf5ac382e8692af8244c5833   [Accessed   22
    January 2020].

35. Jager, J. and Plooy, T. (2010). Information Sources Used To Select A Higher Education
    Institution: Evidence From South African Students, 2 (1), 61-66. Available from
    https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1871187 [Accessed 13 February
    2020].

36. Kabra, R. and Bichkar, R. (2014). STUDENTS' PERFORMANCE PREDICTION
    USING GENETIC ALGORITHM. International Journal of Computer Engineering

and Applications, 5 (3), 19-29. Available from
https://www.semanticscholar.org/paper/STUDENTS%E2%80%99-
PERFORMANCE-PREDICTION-USING-GENETIC-Kabra-
Bichkar/7031e019e8ee13e14e88f67e912256dd19991b50 [Accessed 11 Nov. 2019].

37. Kazi, A. and Akhlaq, A. (2017). Factors Affecting Students' Career Choice. Journal of Research and Reflections in Education, 2, 187-196. Available from https://pdfs.semanticscholar.org/61ca/7b4309f709a3af55aaef8da483fafa6328bf.pdf [Accessed 24 February 2020].

38. Kiss, M., Kotsis, Á. and Kun, A. (2014). The Relationship between Intelligence, Emotional Intelligence, Personality Styles and Academic Success. Business Education & Accreditation, 6 (2), 23-34. Available from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2449352 [Accessed 24 March 2020].

39. Komarraju, M., Ramsey, A. and Rinella, V. (2013). Cognitive and non-cognitive predictors of college readiness and performance: Role of academic discipline. Learning and Individual Differences, 24, 103-109. Available from 10.1016/j.lindif.2012.12.007 https://www.sciencedirect.com/science/article/pii/S1041608012001884 [Accessed 10 October 2019].

40. Lairio, M. and Penttinen, L. (2006). Students' career concerns: challenges facing guidance providers in higher education. International Journal for Educational and Vocational Guidance, 6 (3), 143-157. Available from 10.1007/s10775-006-9107-z https://link.springer.com/article/10.1007/s10775-006-9107-z [Accessed 10 February 2020].

41. MacCann, C., Jiang, Y., Double, K., Bucich, M. and Minbashian, A. (2019). Emotional Intelligence Predicts Academic Performance: A Meta-Analysis. *Psychological Bulletin*, 146 (2), 150-185. Available from https://www.apa.org/pubs/journals/releases/bul-bul0000219.pdf [Accessed 9 February 2020].

42. Mohzan, M., Hassan, N. and Halil, N. (2013). The Influence of Emotional Intelligence on Academic Achievement. Procedia - Social and Behavioral Sciences, 90, 303-312. Available from 10.1016/j.sbspro.2013.07.095

https://www.sciencedirect.com/science/article/pii/S1877042813019678?via%3Dihub [Accessed 3 March 2020].

43. Najimi, A., Sharifirad, G., Amini, M. and Meftagh, S. (2013). Academic failure and students′ viewpoint: The influence of individual, internal and external organizational factors. Journal of Education and Health Promotion, 2 (1), 22. Available from 10.4103/2277-9531.112698 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3778557/ [Accessed 2 September 2019].

44. Olson, R. and Moore, J. (2016). TPOT: A Tree-Based Pipeline Optimization Tool for Automating Machine Learning. JMLR: Workshop and Conference Proceedings, 64, 66-74. Available from https://link.springer.com/chapter/10.1007/978-3-030-05318-5_8 [Accessed 25 November 2019].

45. Ozturk, V. (2013). Selection of appropriate software development life cycle using fuzzy logic. 797-810. Available from 10.3233/IFS-120686 https://dl.acm.org/doi/10.5555/2596266.2596294 [Accessed 8 March 2020]

46. Romero, C. and Ventura, S. (2010). Educational Data Mining: A Review of the State of the Art. IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C, 40 (6), 601-618. Available from https://doi.org/10.1109/TSMCC.2010.2053532 https://ieeexplore.ieee.org/document/5524021 [Accessed 11 February 2020].

47. Pal, V. and Bhatt, V. (2019). Performance Prediction for Post Graduate Students using Artificial Neural Network. International Journal of Innovative Technology and Exploring Engineering, 8 (7), 446-454. Available from https://www.ijitee.org/wp-content/uploads/papers/v8i7s2/G10760587S219.pdf [Accessed 2 November 2019].

48. Sach, R., Petre, M. and Sharp, H. (2010). The use of MBTI in Software Engineering. 1-8. Available from https://www.semanticscholar.org/paper/The-use-of-MBTI-in-Software-Engineering-Sach-Petre/193cdef613c3edbef7ebba96d19bbd9dec773dd5 [Accessed 4 March 2020].

49. Saunders, M., 2009. Understanding Research Philosophies And Approaches. [ebook] Pearson Education, pp.122 - 152. Available at: <https://www.researchgate.net/publication/309102603_Understanding_research_philosophies_and_approaches> [Accessed 6 March 2020].

50. Schlag, S., Schmitt, M. and Schulz, C. (2019). Faster Support Vector Machines. In 2019 Proceedings of the Twenty-First Workshop on Algorithm Engineering and Experiments (ALENEX), pp.199-210. Available at: https://epubs.siam.org/doi/abs/10.1137/1.9781611975499.16 [Accessed 14 Nov. 2019].

51. Singh, A., Thakur, N. and Sharma, A. (2016). A review of supervised machine learning algorithms. 2016 3rd International Conference on Computing for Sustainable Global Development, 1310-1315. Available from https://ieeexplore.ieee.org/abstract/document/7724478 [Accessed 25 January 2020].

52. Singh, P. (2015). Interaction Effect of Brain Hemispheric Dominance and Home Environment on Academic Achievement in Mathematics. International Journal of Science and Research (IJSR), 4 (11), 1940-1943. Available from 10.21275/v4i11.20111502 https://www.semanticscholar.org/paper/Interaction-effect-of-brain-hemispheric-dominance-Singh/a6996559f4d9e7e29c5262ab0826b957bb56ff4b [Accessed 1 March 2020].

53. Tlili, A., Essalmi, F., Jemni, M., Kinshuk and Chen, N. (2016). Role of personality in computer based learning. Computers in Human Behavior, 64, 805-813. Available from 10.1016/j.chb.2016.07.043 https://www.sciencedirect.com/science/article/pii/S0747563216305337 [Accessed 4 February 2020].

54. Tosevski, D., Milovancevic, M. and Gajic, S. (2010). Personality and psychopathology of university students. Current Opinion in Psychiatry, 23 (1), 48-52. Available from 10.1097/yco.0b013e328333d625 https://journals.lww.com/co-psychiatry/Fulltext/2010/01000/Personality_and_psychopathology_of_university.10.aspx [Accessed 4 October 2019].

55. Vavpotic, D. and Bajec, M., 2009. An approach for concurrent evaluation of technical and social aspects of software development methodologies. Information and Software Technology, 51(2), pp.528-545. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0950584908000931 [Accessed 11 February 2020].

56. XIN, Y., KONG, L., CHEN, Y., LI, Y., ZHU, H., GAO, M., HOU, H. and WANG, C. (2018). Machine Learning and Deep Learning Methods for Cybersecurity. 6. Available

from   https://doi.org/10.1109/ACCESS.2018.2836950
https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8359287   [Accessed   14
December 2019].

57. Xu, J., Xing, T. and van der Schaar, M. (2016). Personalized Course Sequence
    Recommendations. IEEE Transactions on Signal Processing, 64 (20), 5340-5352.
    Available   from   10.1109/tsp.2016.2595495
    https://ieeexplore.ieee.org/document/7524023 [Accessed 6 October 2019].

58. Xu, J., Moon, K. and Schaar, M. (2017). A Machine Learning Approach for Tracking
    and Predicting Student Performance in Degree Programs. IEEE Journal of Selected
    Topics   in   Signal   Processing,   11(5),   pp.742-753.   Available   at:
    https://ieeexplore.ieee.org/document/7894238 [Accessed 12 Nov. 2019].

59. Yang, Q., Liu, Y., Chen, T. and Tong, Y., 2019. Federated Machine Learning. ACM
    Transactions on Intelligent Systems and Technology, 10(2), pp.1-19. Available at:
    https://dl.acm.org/doi/10.1145/3298981 [Accessed 10 October 2019].

60. Yao, H., Lian, D., Cao, Y., Wu, Y. and Zhou, T. (2019). Predicting Academic
    Performance for College Students. ACM Transactions on Intelligent Systems and
    Technology,   10   (3),   1-21.   Available   from   10.1145/3299087
    https://dl.acm.org/doi/10.1145/3299087 [Accessed 3 November 2019].

61. Yusoff, M. (2019). Multiple Mini Interview as an admission tool in higher education:
    Insights from a systematic review. Journal of Taibah University Medical Sciences, 14
    (3), 203-240. Available from 10.1016/j.jtumed.2019.03.006
    https://www.sciencedirect.com/science/article/pii/S1658361219300496 [Accessed 7
    February 2020].

62. Zhang, X., Hu, T., Dai, H. and Li, X. (2010). Software Development Methodologies,
    Trends and Implications: A Testing Centric View. Information Technology Journal, 9
    (8), 1747-1753. Available from 10.3923/itj.2010.1747.1753
    https://www.semanticscholar.org/paper/Software-Development-Methodologies%2C-
    Trends-and-A-Zhang-Hu/d2ff1464ccf9cbdf02a290992c7febb26bfb2d65 [Accessed 5
    March 2020].

Bibliography

1. BCS, THE CHARTERED INSTITUTE FOR IT CODE OF CONDUCT FOR BCS MEMBERS (2020). 5th ed. Swindon: BCS, The Chartered Institute for IT. Available from https://www.bcs.org/media/2211/bcs-code-of-conduct.pdf [Accessed 26 June 2020].

2. Carter, P. (2005). *THE COMPLETE BOOK OF INTELLIGENCE TESTS*. Chichester, West Sussex PO19 8SQ, England: Wiley Editorial Offices. Available from https://www.shortcutstv.com/blog/wp-content/uploads/2016/10/sim_IQ-test2.pdf [Accessed 6 December 2019].

3. Cherniss, C. and Goleman, D. (2020). THE EMOTIONALLY INTELLIGENT WORKPLACE. Jossey-Bass. Available from https://www.ftms.edu.my/images/Document/MOD00946%20-%20Organisarional%20transformatoin%20practice/Goleman%20BOOK%20The%20Emotionally%20Intelligent%20Workplace.PDF [Accessed 27 December 2019]

4. Cox, B. (2016). Your Secret Self: Understanding yourself and others using the Myers-Briggs personality test (The MBTI Personality Types Series Book 1). Available from https://www.amazon.com/Your-Secret-Self-Understanding-Myers-Briggs-ebook/dp/B01D3UM9R0 [Accessed 27 July 2020].

5. Freed, J. and Parsons, L. (2012). Right-Brained Children in a Left-Brained World. Riverside: Simon & Schuster.

6. GIBBS, N. (1995). *The EQ Factor New brain research suggests that emotions, not IQ, may be the true measure of human intelligence*. New York: Time Inc. Magazine Company.

7. Hobart, J. and Cano, S. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: the role of new psychometric methods, 1st ed. UK: Health Technology Assessment NIHR HTA Programme. Available from https://pdfs.semanticscholar.org/f021/0a64b4bca4f87629cf7f6c44e8cae18d967e.pdf [Accessed 26 January 2020].

8. Larson, E. and Gray, C. (2020). Project Management, 6th ed. United States: McGraw-Hill Education. Available from https://files.transtutors.com/cdn/uploadassignments/2697846_1_mcgraw-hill-education--2013-.pdf [Accessed 16 March 2020].

9. Mackintosh, N. (2011). IQ and Human Intelligence, 2nd ed. New York: Oxford University Press Inc. Available from http://fdslive.oup.com/www.oup.com/academic/pdf/13/9780199585595.pdf [Accessed 7 January 2020].

10. CARTER, P. and RUSSELL, K. (2020). The Ultimate IQ Test Book. London & Philadelphia: Kogan Page. Available from https://sunsreynat.files.wordpress.com/2014/07/the-ultimate-iq-test-book.pdf [Accessed 18 December 2019].

11. C. Victor Bunderson (Ph.D), The Validity Of The Herrmann Brain Dominance Instrument

APPENDICES

Appendix A: Introduction

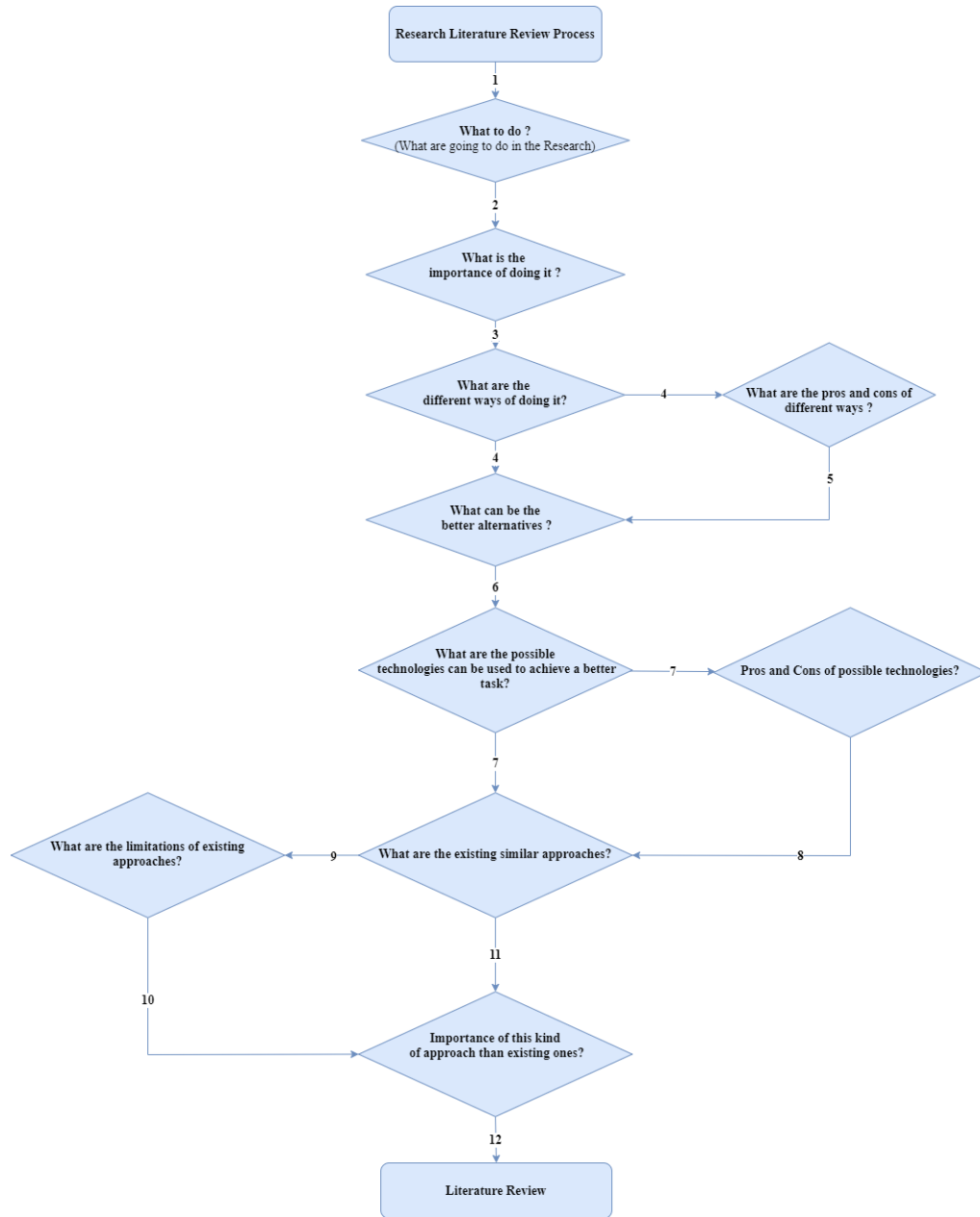## Appendix B: Literature Review

Appendix – B 1 – Concept Graph

Figure B. 1: Concept Graph

Appendix – B 2 - Approaches for seeking opportunities in HE

Approach	Specific Career Goals

Description - Presence of career goals tends school leavers to search and select a specific pathway in HE. This approach can be clustered into four categories (Hull-Blanks et al., 2005).
- School-related goals-he purpose of getting a Bachelor degree in any field
- Job-related goals-targeting professions with a recognised job title (doctor, lawyer)
- Value-related goals – focused more on the value, such as prestige and money.
- Unknown goals - no proper idea about career plans but encouraged to enroll for HE

Impact - Based on personal interest, skills, academic performances and peer influences, students tend to choose a career path for their future (Kazi and Akhlaq, 2018). Proper understanding of both interest career areas and associated HE programs are necessary to make a better decision. Hull-Blanks et al. (2015) further explained that student often select college majors to make parents happy or reduce anxiety related to being undecided and these decisions are made without an understanding of the suitable career for the student. Incorrect career choice causes all resources and individual effort into wrong directions when the academic selections are not aligned with the student's capability.

Approach	General Information Searching

Description -Students who are not having clear career plans, start searching for information on various possibilities of studies before entering HE. Two types of processes can be identified as internal and external information search (Eidimtas and Juceviciene, 2014) from the environment.
- Internal information sources- Student tries to retrieve information from long-term memory based on previously acquired personal experiences and past HE related searches
- Two types of external information sources - Personal sources like family, friends, public opinion leaders, reference groups, peers and independent sources like mass media, internet, university open-door fairs, exhibitions, advertisements of institutions, user groups provide direction to choosing HE.

Impact - Information is gathered from internal and external sources. This approach takes a long time in the student's decision-making process in HE. Eidimtas and Juceviciene (2014) explained if school-leavers who have no previous experience and choosing study

for the first time, are usually involved in the field of external search. That is very large and cannot easily distinguish relevant information and can cause difficulties in choosing the right HE path. Jager and Plooy (2010) provide evidence that the opinion of parents and family members has a strong involvement in the decision of the students. Further, they explained today the choice of the HE institution and course is more complicated than in the past as more alternatives are widely available. Due to this school leaver are being bombarded with different types of commercials and information on HE institutions and courses. However, the wide availability of information allows prospective students to make informed choices, and it is essential to consider the quality of the course or career success based on the students' ability before making a decision.

Approach	**Academic guidance and counselling**
Description - Two different methods are available in this approach as study choice guidance and student counselling. Counsellors support school-leavers when structuring a pathway in HE and help parents in understanding their role in inspiration and motivation of the children by forming their expectations (Eidimtas and Juceviciene, 2014). Academic and career counsellors help school leavers to build relationships between academic integrity and career goals. Hull-Blanks et al. (2005) explained professional counselling services and university counselling centers are especially helpful because of the opportunity for one-to-one interaction with students or online counselling.	
Impact - Student guidance and counselling can be considered as proactive support for school leavers educational, vocational and personal development respond to the changing needs of individual students rather than as a remedy to problem situations (Lairio and Penttinen, 2006). Educational guidance is important when a student is selecting a HE program. Student counsellor professionals are responsible for accessing the latest information on employment trends, variety of requirements and collaborate to mentor culture, active networking with subject departments, and guiding students to a suitable pathway.	

Table B. 1: Approaches for seeking opportunities in HE

Appendix – B 3 - Student guidance and support techniques in HE

Technique 1	University Entrance Exams (Aptitude Tests)
Description	Completion of college entrance exams also called college readiness is required to enter some universities and programs. Aptitude tests are designed to measure students' cognitive abilities as verbal and numerical skills to predict academic

	performances (Walsh, 2013, p390). Students are guided for a program according to the score achieved from the test.

Advantages	Disadvantages
1. More effective than many other strategies for measuring attitudes, such as oral exams and the learning portfolios. 2. A standardised aptitude test (i.e., SAT) can be used as a quantitative metric which universities can use to compare students from different areas.	1. Aptitude tests don't measure student's non-cognitive abilities (personality, emotions, attitudes) 2. Psychological effects during the test such as 'test anxiety' can be negatively effect on test performances

Technique 2	Interviews
Description	Interviews for the selection of students is an important method, especially in health and medical science fields in HE institutions seeking capable students to enroll programs (Yusoff, 2019).

Advantage	Disadvantage
1. Effective way to ensure candidate's communication skills 2. Student get a chance to demonstrate own skills	1. A time and resource consuming process 2. May cause to excitement and nervousness in student

Table B. 2 Student guidance and support techniques in HE

Appendix – B 4 - Comparison of Personality Indicators

MBTI	FFM	MMPI
This Model was developed by Katharine and Isabel Briggs during 1942 - 1944 based on Carl G. Jung's approach on Psychological types (Tlili et al., 2016) as a support for the spies of World War II aimed to compare specific tasks. Based on preferences for Energy, Information, Decision and Lifestyle of individuals 4 types of	The Five-Factor Model was developed from 1980 onwards. This model was not built with a particular theoretical perspective but (Tlili et al., 2016) used some lexical hypothesis developed based on individuals' opinion on others and themselves. Five dimensions have been introduced as 1. Openness to experience 2. Conscientiousness 3. Extraversion	MMPI was developed around 1930 at the University of Minnesota. Versioning of tests has been introduced as MMPI-1 and MMPI-2. These tests are commonly used by mental health professionals. 10 scales of the MMPI have been defined as 1.Hypochondriasis 2. Depression 3. Hysteria 4. Psychopathic Deviate 5. Masculinity -femininity 6. Paranoia

dichotomies are explained in this Model. 1. Extraversion(E) Introversion(I) 2. Sensing (S) - Intuition 3. Thinking (T) - Feeling (F) 4. Judging (J) or Perceiving (P) 16 personality tests widely used by employers to identify employee strengths and weaknesses	4. Agreeableness 5. Neuroticism NEO personality inventories have been developed to identify FFM personalities.	7. Psychasthenia 8. Schizophrenia 9. Hypomania 10. Social introversion.

Table B. 3 Comparison of Personality Indicators

Appendix – B 5 - Validity of the selected Psychometric Test in Plato

Test Type	Personality	IQ	EQ	Brain Dominance
Source	Cox, B. (2016). *Your Secret Self: Understanding yourself and others using the Myers-Briggs personality test (The MBTI Personality Types Series Book 1).* Available from https://www.amazon.com/Your-Secret-Self-Understanding-Myers-Briggs-ebook/dp/B01D3UM9R0 [Accessed 27 December 2019].	CARTER, P. and RUSSELL, K. (2020). *The Ultimate IQ Test Book.* London & Philadelphia: Kogan Page. Available from https://sunsreynat.files.wordpress.com/2014/07/the-ultimate-iq-test-book.pdf [Accessed 18 December 2019].	GIBBS, N. (1995). *The EQ Factor New brain research suggests that emotions, not IQ, may be the true measure of human intelligence.* New York: Time Inc. Magazine Company.	Freed, J. and Parsons, L. (2012). *Right-Brained Children in a Left-Brained World.* Riverside: Simon & Schuster.

Table B. 4: Validity of the selected Psychometric Tests in Plato

Appendix – B 6 - Classification algorithms

Below table includes a comparison of a few classification type ML algorithms by considering their pros and cons to heaven idea about the behaviour of classification algorithms.

Naive Bayes Classifier
Pros: Good in filtering (Granik and Mesyura, 2017), Memory efficient, Simplicity high (Gupte et al.,2014), Good for real-time prediction   **Cons**: It may be prone to overfit. Usually, identify as bad estimators.
Support Vector Machines
Pros: Comparatively memory efficient, Less impact from outliers in the dataset.   **Cons**: Usually not good in large dataset training because of its time complexity. (Schlag, Schmidt and Schulz, 2019) When classes are overlapped, performance is dropping.
DTs: (Widely used in the education domain for prediction tasks)
Pros: More interpretable since it provides readable rule sets (Ali et al., 2012) For generating a model it needs low computational cost (Kabra and Bichkar, 2014)   **Cons**: Can easily overfit, Split nodes sometimes become costly (Ali et al., 2012), Can lead to suboptimal solutions (Kabra and Bichkar, 2014)
Random Forest
Pros: Able to overcome overfitting issues, accuracy and importance of variables generated automatically. Simplicity is high, High Accuracy (Gupte et al., 2014) Perform well in imbalanced datasets. Can handle a large amount of data Bagging type ensemble technique is used in RF   **Cons**: More complex than DTs. The prediction process is time costly. Memory consumption is High. (Gupte et al.,2014)
ANNs
Pros: Very efficient and accurate with large data sets.   **Cons**: Need a comparatively large data set for accurate prediction. Not designed for a small data set. The model Training process is not interpretable (Gayathri and Marimuth, 2016).

Table B. 5: Comparison of a few classification type ML algorithms

Appendix – B 7 - Regression algorithms

Regression algorithms are SML types which are used to predict continuous values (i.e. a number). Below table contains the pros and cons of some regression algorithms

Linear regression
Pros: Most simple regression types. Less computational power.   **Cons**: Can be used only to identify the relationship between 2 continuous variables. (linear models)
Lasso regression

Pros: Can be used for automatic feature selection for regression models. Comparatively fast and able to avoid overfitting of the model. **Cons**: Created models are less interpretable. Sometimes models can be not useful since models are unstable (heavily dependent on data) usually MSE is high.	
Support vector regression **Pros**: The computational complexity of SVM is independent of the input dimensionality. Good generalisation capability (Awad and Khanna, 2015) **Cons**: Not perform well in overlapped data. Performance decreases when the data set is too large. Prone to over fit if the noise of the data is high	
Random forest regression **Pros**: Perform well in imbalance data sets. Since ensemble techniques are used, accuracy is high. This is a bagging type ensemble algorithm which contains multiple DTs. **Cons**: High computational power is required	

Table B. 6: Pros and cons of some regression algorithms

Appendix C: Methodology

Appendix C – 1 - Software Development Methodologies

Methodology	Waterfall Model	
Characteristics	Linear sequential development life cycle with requirement, design, implementation, verification and maintenance phases (Zhang et al., 2010)	
Pros 1.Suitable for small projects with clear requirements 2. Features realisation is easy to test and analyse 3. Each phase has specific deliverables 4.Easy to use and understand phases with stable product definitions		**Cons** 1. Requirement changes are not accommodated, Rigidity model 3.Not suitable for maintenance projects 4.Doesn't handle project risks
Methodology	Agile (Scrum)	
Characteristics	An iterative model with planning, requirement analysis, design, coding, unit and acceptance testing phases. (Zhang et al., 2010)	
Pros 1. Products are delivered in a short time 2. A large team can effectively manage 3. An adaptive approach for requirement changes		**Cons** 1 .Difficult to handle complex dependencies 2.Depends heavily on customer interactions 3.Not suitable for inexperienced developers

Methodology	Spiral Model
Characteristics	An iterative and incremental model with 4 phases (Zhang et al., 2010) 1.) Planning 2.) Risk Analysis 3.) Engineering 4.) Evaluation The final system is constructed based on the refined prototypes

Pros	Cons
1. Suitable for medium to large size projects. 2. Development can be divided into smaller parts and requirement changes are accepted 3. A high amount of risk analysis is involved and avoids possible risks. 4. Extensive use of prototypes and strong documentation 5. Additional functionalities can be added at later and customised products can be developed.	1. Cost involvement for this model is high 2. Spiral may continue indefinitely 3. Project success depends on the risk analysis phase 4. Domain experts are required for risk analysis

Methodology	RAD – Rapid Application Development Model
Characteristics	A model with minimum planning and focus prototype development including business modelling, data modelling, process modelling, application generation, testing and turnover phases.

Pros	Cons
1. Reduces development time and quick customer review 2. Code can be reused 3. Integration issues solve from the very beginning and requirements are accommodated	1.Depends on the strong design and business team, Weak documentation 2.User involvement throughout the development life cycle and highly depends on modelling skills

Methodology	Prototyping Model
Characteristics	The developer creates prototypes of solutions before development. Includes requirement identification, prototyping, review and engineering phases.

Pros	Cons
1.Possible to get customer feedback earlier 2. The exact requirement can be identified 3.Reduces time as defects can be identified earlier and risk can be identified early	1. Too much client involvements 2. A large number of modifications can be done and it is not good for a successful project 3. Scope of the project may expand beyond the plan 4. More programming effort is needed if prototype development may not be properly monitored.

Table C. 1: Comparison of Software development methodologies

Appendix D: Project Management
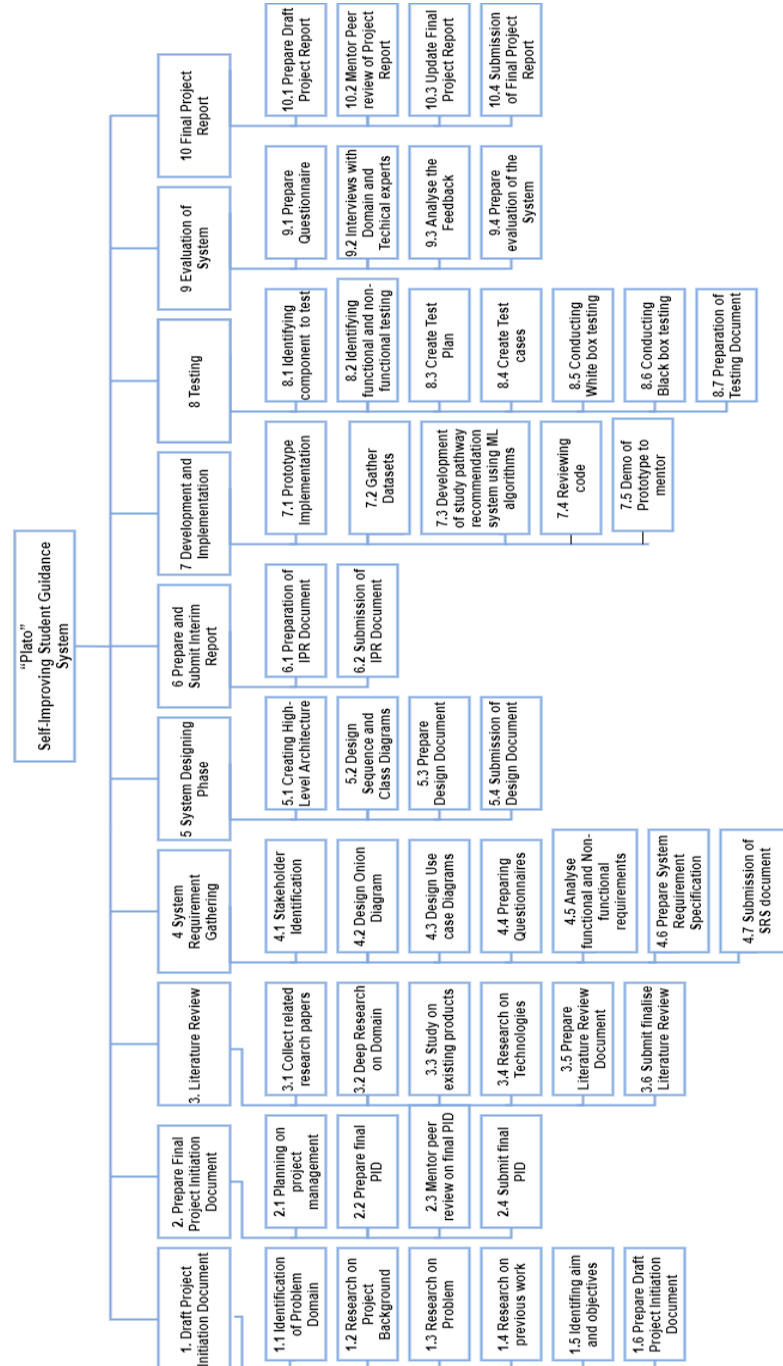
Appendix D1 - Work Breakdown Structure of Project

Figure D. 1: Work Breakdown Structure of Project

Appendix D2-Detailed Activity Schedule

WBS	Task Name	Start Date	End Date	Duration (Days)
1	**Draft Project Initiation Document**	**8/4/2019**	**9/23/2019**	**50**
1.1	Identification of Problem Domain	8/4/2019	8/20/2019	16
1.2	Research on Project Background	8/20/2019	9/2/2019	13
1.3	Research on Problem	9/3/2019	9/10/2019	7
1.4	Research on previous work	9/10/2019	9/16/2019	6
1.5	Identifying aim and objectives	9/17/2019	9/20/2019	3
1.6	Prepare Draft Project Initiation Document	9/20/2019	9/23/2019	3
2	**Prepare Final Project Initiation Document**	**9/23/2019**	**10/20/2019**	**27**
2.1	Planning on project management	9/25/2019	9/30/2019	5
2.2	Prepare final PID	10/1/2019	10/18/2019	17
2.3	Mentor peer review on final PID	10/18/2019	10/19/2019	1
2.4	Submit final PID	10/20/2019	10/20/2019	0
3	**Literature Review**	**8/10/2019**	**11/14/2019**	**96**
3.1	Collect related research papers	8/10/2019	8/20/2019	10
3.2	Research on Domain	8/21/2019	9/30/2019	40
3.3	Study on existing products	10/21/2019	11/1/2019	11
3.4	Research on Technologies	11/2/2019	11/12/2019	10
3.5	Prepare Literature Review Document	10/30/2019	11/14/2019	15
3.6	Submit Finalise Literature Review	11/14/2019	11/14/2019	0
4	**System Requirement Gathering**	**11/20/2019**	**1/14/2020**	**55**
4.1	Stakeholder Identification	11/20/2019	11/25/2019	5
4.2	Design Onion Diagram	11/25/2019	11/26/2019	1
4.3	Design Use Case Diagrams	11/27/2019	12/7/2019	10
4.4	Preparing Questionnaires	12/20/2019	12/22/2019	2
4.5	Analyse functional and Non-functional requirements	12/23/2019	12/28/2019	5
4.6	Prepare System Requirement Specification	12/20/2019	1/12/2020	23
4.7	Submission of SRS document	1/14/2020	1/14/2020	0
5	**System Designing Phase**	**11/30/2019**	**1/14/2020**	**45**
5.1	Creating High- Level Architecture	12/1/2019	12/3/2019	2
5.2	Design Sequence and Class Diagrams	12/4/2019	12/14/2019	10
5.3	Prepare Design Document	12/15/2019	12/30/2019	15
5.4	Submission of Design Document	11/25/2019	1/14/2020	50
6	**Prepare and Submit Interim Report**	**1/20/2020**	**2/10/2020**	**21**
6.1	Preparation of IPR Document	1/20/2020	2/10/2020	21
6.2	Submission of IPR Document	2/10/2020	2/10/2020	0

7	**Development and Implementation**	**1/25/2020**	**3/30/2020**	**65**
7.1	Design and host Plato test page	1/20/2020	2/15/2020	26
7.2	Algorithm Analysis	2/16/2020	2/24/2020	8
7.3	Development of core functionalities	2/25/2020	3/16/2020	20
7.4	Implementation of application of the Plato system	3/2/2020	3/19/2020	17
7.5	Reviewing code	3/21/2020	3/30/2020	9
7.6	Demo of Prototype to mentor	4/2/2020	4/2/2020	0
8	**Testing**	**2/28/2020**	**4/20/2020**	**52**
8.1	Identifying component to test	2/28/2020	3/1/2020	2
8.2	Identifying functional and nonfunctional testing	3/1/2020	3/3/2020	2
8.3	Create Test Plan	3/4/2020	3/12/2020	8
8.4	Create Test cases	3/13/2020	3/20/2020	7
8.5	Conducting White box testing	3/20/2020	4/15/2020	26
8.6	Conducting Black box testing	3/20/2020	4/15/2020	26
8.7	Preparation of Testing Document	3/5/2020	4/20/2020	46
	Unable to continue the project due to Sickness	18/4/2020	31/5/2020	43
	Submitted a mitigating circumstances for Deferral	18/4/2020	23/4/2020	5
9	**Evaluation of System**	5/6/2020	20/7/2020	45
9.1	Prepare Questionnaire	5/6/2020	6/6/2020	1
9.2	Interviews with Domain and Technical experts	7/6/2020	14/7/2020	37
9.3	Analyse the Feedback	7/7/2020	20/7/2020	13
9.4	Prepare evaluation of the System	5/6/2020	20/7/2020	45
10	**Final Project Report**	1/6/2020	27/7/2020	56
10.1	Prepare Draft Project Report	1/6/2020	1/7/2020	30
10.2	Mentor Peer review of Project Report	5/7/2020	8/7/2020	3
10.3	Update Final Project Report	10/7/2020	26/7/2020	16
10.4	Submission of Final Project Report	27/7/2020	27/7/2020	0

Table D. 1: Detailed Activity Schedule

Appendix D3 - Gantt chart
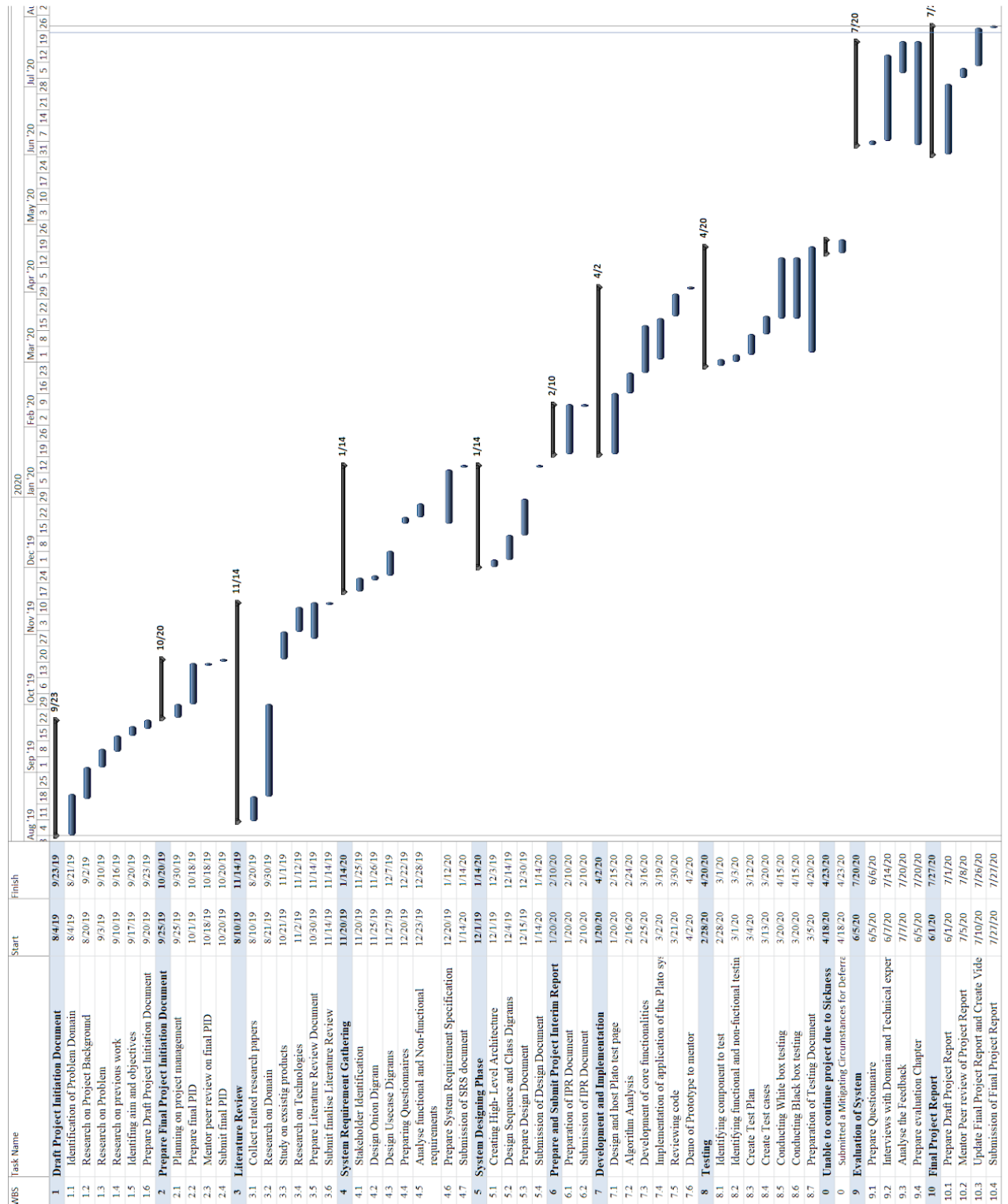
Figure D. 2: Gantt chart

Appendix E: Software Requirement Specification

Appendix E 1- Survey Questioner Form

Figure E. 1: Survey Questioner Form

Appendix F: Testing

Appendix G: Evaluation

PLATO – Project Evaluation

I am Sapumal Perera, a BEng (Hons) Software Engineering undergraduate at Informatics Institute of Technology (IIT) affiliated with the University of Westminster, UK.

Thank you for taking your valuable time to provide an evaluation for my project. This questionnaire is distributed as apart of my final year research project which will help me a lot in preparing my bachelors thesis. Please note that this feedback is for completion of my undergraduate research project and will not be used for any other purpose.

I would appreciate it if you could take some time to provide your thoughts and feedback on the project after watching the PLATO demonstration video and the overview documentation.

PLATO Demonstration Video - https://youtu.be/4Ru833o8inM
Overview Documentation - http://platoguide.s3-website-us-west-2.amazonaws.com/Personality.html

Thanks in advance!

*Required

Email address *

Your email address

Introduction to Plato

PLATO

| Plato | psychometric-...

Student Guidance System
Helps you to put a ding in the universe by revealing your talents.

The direction in which education starts a man will determine his future in life
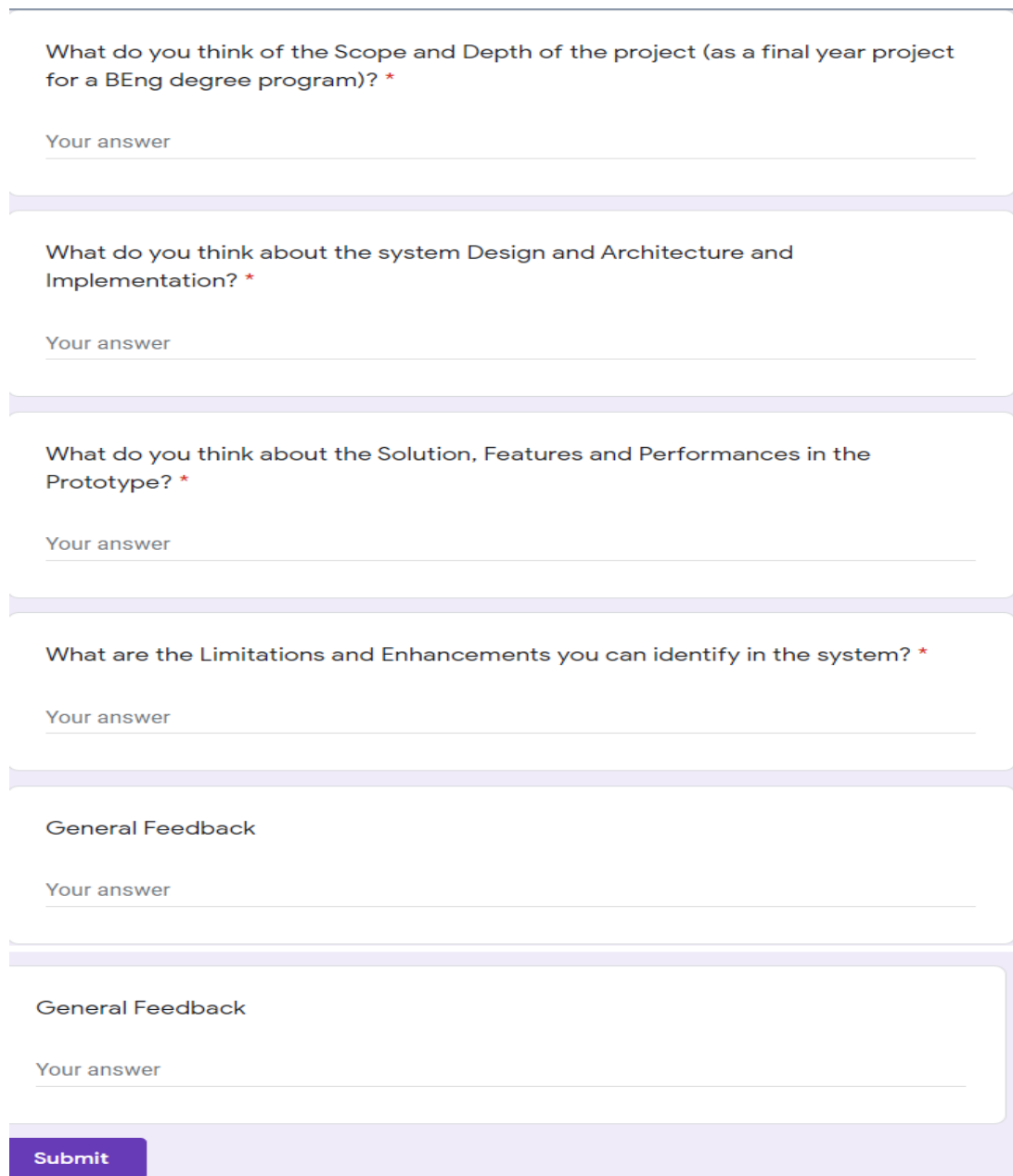
— Plato —

Your Name *

Your answer

Designation or a small Description about you *

Your answer

What do you think of the overall concept of the Project? *

Your answer

Figure G. 1: Evaluation Form