# CLASSIFICATION OF ONLINE INSTRUCTIONAL STRATEGIES BASED ON PERSONALISATION


**WILLIAM HENG CHUN MENG**

**FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**
**UNIVERSITY OF MALAYA**
**KUALA LUMPUR**


**2022**

# CLASSIFICATION OF ONLINE INSTRUCTIONAL STRATEGIES BASED ON PERSONALISATION

**WILLIAM HENG CHUN MENG**

**RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF DATA SCIENCE**

**FACULTY OF COMPUTER SCIENCE AND INFORMATION TECHNOLOGY**
**UNIVERSITY OF MALAYA**
**KUALA LUMPUR**

**2022**

**UNIVERSITY OF MALAYA**

**ORIGINAL LITERARY WORK DECLARATION**

Name of Candidate: WILLIAM HENG CHUN MENG (I.C No: 880310-56-5161)

Matric No: S2005592

Name of Degree: MASTER OF DATA SCIENCE

Title of Project Paper/Research Report/Dissertation/Thesis (“this Work”):

CLASSIFICATION OF ONLINE INSTRUCTIONAL STRATEGIES BASED ON PERSONALISATION

Field of Study: Machine Learning

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya (“UM”), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate’s Signature Date: 23 June 2022

Subscribed and solemnly declared before,

Witness’s Signature

ASSOC. PROF. DR. NOR LIYANA MOHD SHUIB
DEPARTMENT OF INFORMATION SYSTEMS
FACULTY OF COMPUTER SCIENCE &
INFORMATION TECHNOLOGY
UNIVERSITY OF MALAYA
50603 KUALA LUMPUR

Date: 23 June 2022

Name: Dr. Nor Liyana Binti Mohd Shuib

Designation: Assoc Prof

# CLASSIFICATION OF ONLINE INSTRUCTIONAL STRATEGIES BASED ON PERSONALIZATION


## ABSTRACT

There is no one size fits all strategy for most of the things in our life, at least not for education. Personalisation is essential for the learning process as it will make what is learnt by the learner relevant and thus less wasted time and increase the effectiveness of learning. Instruction strategy is the strategy that steers learners to be independent and become strategic learners at the same time to meet their learning objectives. There is plenty of instructional design models available but lacking when it comes to the classification of online instructional strategy based on personalisation, as most of the previous work done were either omitted personalisation or did not deal with classification problems in the first place. Moreover, the advent of COVID-19 led to a meteoric surge in the popularity of online learning. To fill the gap, this study will propose a machine learning technique to perform classification of instructional strategy with the consideration of learner feedback gathered by a survey. Three machine learning algorithms, namely Gradient Boosting, Naïve Bayes and Support Vector Machine, are used to achieve that. The findings show that this study's deliverable would provide useful insight to not just educators and learners but the whole education sector as well.



**Keywords**: Online Instructional Strategies; personalisation; machine learning

**KLASSIFIKASI STRATEGI PENGAJARAN DALAM TALIAN BERDASARKAN KEPERIBADIAN**


## ABSTRAK

Tidak ada satu saiz yang sesuai untuk semua strategi untuk kebanyakan perkara dalam hidup kita, sekurang-kurangnya bukan untuk pendidikan. Pemperibadian adalah penting untuk proses pembelajaran kerana ia akan menjadikan apa yang dipelajari oleh pelajar relevan dan dengan itu mengurangkan masa yang terbuang dan meningkatkan keberkesanan pembelajaran. Strategi pengajaran ialah strategi yang mengarahkan pelajar untuk berdikari dan menjadi pelajar strategik pada masa yang sama untuk mencapai objektif pembelajaran mereka. Terdapat banyak model reka bentuk pengajaran yang tersedia tetapi kurang apabila ia datang kepada klasifikasi strategi pengajaran dalam talian berdasarkan pemperibadian, kerana kebanyakan kerja yang dilakukan sebelum ini sama ada ditinggalkan pemperibadian atau tidak menangani masalah klasifikasi pada mulanya. Selain itu, kemunculan COVID-19 membawa kepada lonjakan mendadak dalam populariti pembelajaran dalam talian. Untuk mengisi jurang, kajian ini akan mencadangkan pendekatan pembelajaran mesin untuk melaksanakan klasifikasi strategi pengajaran dengan mengambil kira maklum balas pelajar yang dikumpul melalui tinjauan. Tiga algoritma pembelajaran mesin, iaitu Gradient Boosting, Naïve Bayes dan Support Vector Machine, digunakan untuk mencapainya. Dapatan kajian menunjukkan bahawa hasil kajian ini akan memberikan gambaran yang berguna kepada bukan sahaja pendidik dan pelajar tetapi seluruh sektor pendidikan juga.



**Kata Kunci:** Strategi pengajaran dalam talian; keperibadian; pembelajaran mesin

## ACKNOWLEDGEMENTS

First and foremost, I want to give God the Almighty thanks and honour, who has bestowed abundance blessing, wisdom, and opportunity to me. I would not be able to accomplish this research project without His Holy guidance.

There are countless of people supported my effort on course of this Work and I can only name some of them exhaustively. I am deeply indebted to my supervisor Assoc Prof. Dr. Nor Liyana Binti Mohd Shuib of the Faculty of Computer Science and Information Technology at University of Malaya who has been an ideal mentor, and this Work supervisor, offering advice and encouragement with a perfect blend of insight and kindness. Her generosity and support on dataset drive this research project to faultless direction. I'm proud of, and grateful for, my time working with her. She provided invaluable feedback on my analysis and framing, at times responding to emails late at night and early in the morning.

Moreover, I would like to express my profound appreciation and gratitude to my family and partner Wong Mei Chee for their patience and understanding as well as unconditional love. It would be difficult to find adequate words to convey how much I owe them

Lots of love and thank to all of you.

## TABLE OF CONTENTS

# LIST OF FIGURES

## LIST OF TABLES

## LIST OF SYMBOLS AND ABBREVIATIONS

| | | |
|---|---|---|
| GB | : | Gradient Boosting |
| NB | : | Naïve Bayes |
| SVM | : | Support Vector Machine |
| AUC | : | Area Under the Curve |
| ROC | : | Receiver Operating Characteristic |

# CHAPTER 1: INTRODUCTION

This chapter introduces the research background, problem, objectives, questions, method, scope and significance of classification of instructional strategies based on personalisation.

## 1.1 Research Background

The meteoric surge in popularity of online learning, particularly since the COVID-19 pandemic, is rapidly transforming the way students acquire knowledge and certification (Dhawan, 2020). Modern days technologies not just make online learning possible but also promoted a high degree of personalisation for almost everything and matter in our daily life, including learning style and strategy. Kim et al. (2015) discover that it is beneficial to the student by mere awareness of their own preference of learning style. Each student learns in their own unique style and at their own speed. However, their study programs put them on track to meet the school teaching standard (Johnston, 1994). The evolution of teaching concepts attributes to a shift in teaching thought brought about by the advancement of network technology (Yang, 2022). In order to improve students' online learning experiences, it is critical to establish a classification model of online instructional strategies by examining the concept of student personalisation. It goes against the grain of most institutions' "one size fits all" strategy (Popescu, 2010). Besides, students can become more efficient and effective learners if they are able to develop a personal learning plan by understanding how they learn best based on their strengths, needs, and abilities (McTighe & O'Connor, 2005). In addition, Estrela et al. (2017) argue that keeping the learner engaged during the entire online learning process is particularly very important because the chances are very high that the student probably has lost his focus that, eventually contributes to his performance drop.

## 1.2 Problem Statement

It is of utmost importance to consider student feedback to develop instructional strategies based on personalisation; as Sadler (2014) wrote a paper describing that the feedback of students should help them to understand more about their own learning goals and thus, feedback amplifies learning. There are past research and papers that discuss about designing or developing instructional strategies. However, there is no classification model for instructional strategies based on personalisation (Akdeniz, 2016) other than the design model (Suartama et al., 2019). Furthermore, Yang (2022) recounts that the process of learning cannot be modelled with any existing algorithm by neglecting students' personalisation. That being so, personalisation is extremely invaluable not just for students but also for educators.

Meanwhile, it is very common that instructional strategies are designed to meet measured criteria (van Geel, 2019) rather than based on personalisation (Drljača et al., 2017). We know that developing differentiated instructional strategies can be hard (Gregory & Chapman, 2012) as educators do not have a scientific way but only rely on their own or shared experiences and examples (Cheng et al., 2021).

## 1.3 Research Objectives and Research Questions

In this study, we are going to explore machine learning techniques to perform the classification of online instructional strategies based on students' personalisation. This study seeks to address the research objectives listed below:

Objective 1: To develop a classification model of online instructional strategies based on personalisation.

Objective 2: To evaluate the model performance of online instructional strategies based on personalisation.

The followings are the research questions corresponding to the objectives that we want to address:

Question 1: How should the classification model of online instruction strategies based on personalisation be developed?

Question 2: What is the performance of the classification model that is developed?

## 1.4 Research Scope

This study looks into the use of the machine learning techniques to classify online instructional strategies based on personalisation. The scope of this research work will be as follows:

- Undergraduate and postgraduate students are the target group selected for the purpose of achieving the objective of this study. Thus, materials intended for this research would be limited to university students only.
- Intensive data cleaning is critical to ensure a high level of dataset quality.
- Three machine learning algorithms are employed to perform classification in this research project, namely Gradient Boost (GB), Naïve Bayes (NB) and Support Vector Machine (SVM).

## 1.5 Research Significant

There is no classification model based on personalisation for online instructional strategy currently; hence the significance of this study would be providing a machine learning technique for the educator to develop online instructional strategies and for learners to get recommended online instructional strategy based on their personalisation.

## 1.6 Organisation of Research

This study will be structured into five major chapters:

| | |
|---|---|
| Chapter 1 | The opening chapter begins with a research background that will serve as the introduction and overview of this project. Then it will be followed by a problem statement addressing the domain problems. After that would be a listing of research objectives and research questions before proceeding to clearly itemise every research method. The discussion of limitations as well as the significance of the study take place in this chapter too. Next, the specifics of organisation of research will wrap up the chapter. |
| Chapter 2 | This chapter mainly focuses on examining the literature review for learning, instructional strategies, personalisation, classification model, as well as various related approaches that were being implemented in previous research. A literature map will then be presented to serve as the graphical metaphor in this chapter to organise information for this research project. |
| Chapter 3 | This chapter will deliberate the research methodologies to be employed to achieve each of the project's objectives. It begins with a framework of our research design that outlines the indication of how the study is being conducted. Subsequently, a detailed description of how data is being collected and cleaned will be identified here as well. Finally, will be the discussion about the development of models to realise the research objectives. |
| Chapter 4 | This chapter will then analyse the outcome of our model and discuss about the findings derived in connection with the research objectives. After that would be evaluation of model and interpret the analysis. Lastly, exploration of potential insights can be highlighted in the findings consequentially. |

| | |
|---|---|
| Chapter 5 | This chapter will serve as the conclusion of the study by providing an overview of what was done to attain its objectives. Afterwards, the research project’s contribution, limitations and also recommendations for future work will be discussed in this chapter. |

# CHAPTER 2: LITERATURE REVIEW

## 2.1 Introduction

The central purpose of this chapter is to introduce those important terms that will be used throughout the whole research project. This chapter is essentially broken into four major sections. The first section elucidates what is learning as it is the main domain of our study. Next, the second section discusses student personalisation when it comes to learning. Then the third section delves into prior work in using machine learning techniques to predict instructional strategies as well as narration of their applications limitations. The final section of this chapter will depict the study's frameworks.

## 2.2 Learning

Learning is a transformative process that will change a person's disposition, knowledge, capabilities, or behaviour by acquiring information through collective experience from five basic senses, namely smell, sight, touch, hearing and taste. Learning is a process of receiving information from different external sources and transforming it through internalisation, resulting in internal sources for knowledge. The internal transformation can be pondering, brute or technical memorisation, reflection or inference. For a person to be called a learner, the person will have a continuous learning process to gain knowledge. In this research project, the use of the term student and learner interchangeably, whereas the party who facilitates or guide the learning process as teacher or educator interchangeably as well. It is understandable that every student has different approaches to learn to obtain knowledge; however, not every student knows themselves well enough to maximise their learning potential and performance.

According to Ambrose et al. (2010), there are seven principles of learning when it comes to education. The first principle is prior knowledge, in which what the student knew prior hand could help or hinder learning. The teaching implication of this principle is to discover students' prior knowledge and whether students have self-assess their own familiarity with prior knowledge. After that would be identification and address of errors or student's misconceptions. The second principle is knowledge organisation, in which how students organise their knowledge will influence how they are going to learn as well as apply what they learnt. Teachers can help students to organise what they know and provide useful connections between meaningful and proper principles and concepts. Motivation is the third principle that drives the student to learn and, more importantly, sustain the learning process. For teachers to have an impact on students with regard to motivation would be to show enthusiasm and giving reward to students for outcomes achieved or explain the relevance of what they are learning to real world applications. Basically, creating an environment that is supportive to the student learning process and this will also lay foundation for alignment of outcomes with activities and assessments.

The fourth principle is mastery, it is a measure of how well students handle the knowledge, integrate them internally and practice them externally. For a student to develop mastery, they will have to train themselves repeatedly to know when and what to apply the knowledge. The educator can break down a complex task into smaller component to ease student on learning; this also simultaneously exhibit examples to follow that aid integration. Next, the fifth principle is goal directed; learning with purpose is meaningful, and it improves learning quality when paired with targeted feedback. Teachers are essentially the ideal role in providing invaluable guidance by making reasonable objectives after accessing performance levels and feedback. The sixth principle is intellectual, emotional and social climate, as this plays a pivotal role in a student's course of learning. By fostering a healthy and safe environment to allow students to express their point of view and raise questions will encourage better learning for sure. Now, this is where personalisation kicks in; an open environment with appropriate rules set by teacher that can help develop nourishing interactions. The last principle is self-directed; a student must learn to adjust and adapt their learning independently eventually. Cultivation of self-reflection and developing own metacognition by the teacher will encourage students to model their learning strategies and rubrics so that their learning process will not halt by changing various factors.

### 2.2.1 Instructional Design

The term instructional design refers to an organised method of developing education that is based on adult learning concepts. It makes it easier for learners to acquire information, skills, and attitudes. Instructional design analyses learning objectives and targets before developing a delivery approach to suit those demands. It entails creating instructional strategies and activities, as well as testing and evaluating all instructional and learner activities.

Instructional design is a process that which educational facilitators sketch out the architecture and flow of their education courses by making sure that all directions and parts are included leading to their educational key performance indicators are met.

According to Cennamo & Kalk (2019), there are six characteristics that instructional design models should embodied:

1. Instructional design should focus on the learners and their performance.
2. Clear and precise goals are necessary.
3. Takes real-world performance into consideration thus practical.
4. Outcomes can be measured reliably and reasonably.
5. Emphasise on data which is crucial to the process despite empirical design.
6. Promotes teamwork and value team effort.

#### 2.2.1.1 Instructional Design Model

There are many instructional design models available support instructional design process by providing guidelines to plan suitable pedagogical scenarios to achieve instructional goals that facilitate learning more effectively.

*1) ADDIE Model*

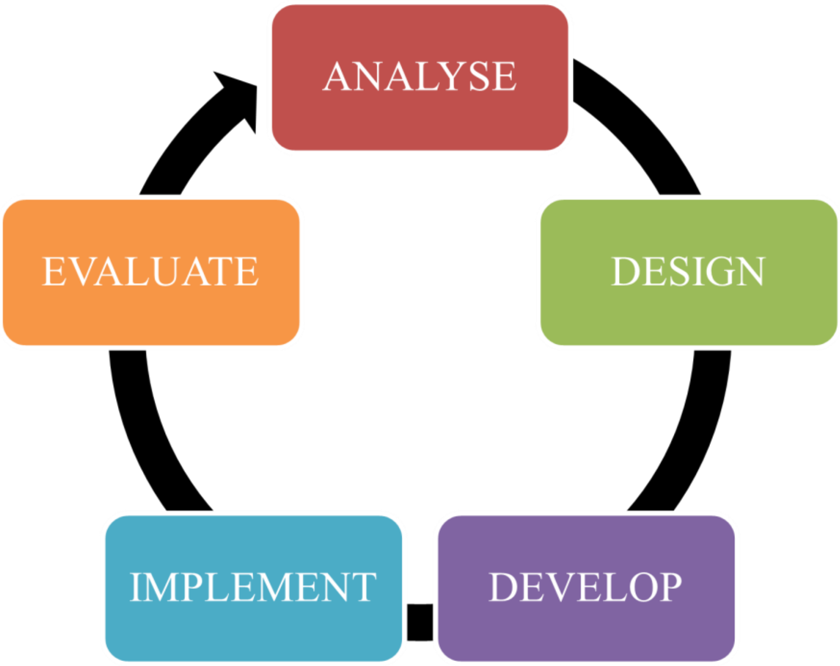


Figure 2.1: ADDIE Model (Molenda, 2003)

Many instructional designers and training developers use the Addie instructional systems design framework to create courses. As a result, the ADDIE approach is being used in a wide array of current instructional system design models.

Each phase in the ADDIE paradigm has a result that feeds into the next step. Analyse, Design, Develop, Implement, and Evaluate is the acronym for the ADDIE model. These acronyms translate into a five phases procedure for the creation of educational substances.

Analyse: The model begins with the instructional designer illuminating the issue or problem that will be addressed with an instructional involvement, denotes the necessity of train, and executing a thorough target crowd analysis to determine the instructional platform, what knowledge had been learnt in prior, abilities and skills, opportunities, and blockers.

Design: Then the designer for instructional will create learning goals and make a decision to determine which instructional tactics will be used to attain them. The look, feel, and operation of the instructional materials, as well as how they are provided to the student, are all deciding factors to decision to be made. This is when making storyboards and e-learning prototypes shines.

Develop: To create instructional strategy or material that supports, content is compiled and incorporated into the design so that learning objectives can be met. The quality of the deliverable will then be checked and revised extensively in this step before moving on to the next step.

Implementation: Upon finishing designing, the completed lesson or the learning object in other words is distributed to the targeted crowd and the effectiveness as well as impact of those will be measured after rolling out.

Evaluate: Lastly, the instructional designer will then employ a variety of techniques to decide whether the instructional strategy or the learning material is producing the desired outcomes that deliver the expected result to meet the learning objectives set earlier on.

*2) ASSURE Model*

ASSURE is another instructional design model with the tenets to provide improvement for teaching and increasing effectiveness for the learning process. The abbreviation "ASSURE" stands for the five different procedures designated in the model. Every stage is meticulously described as follows:

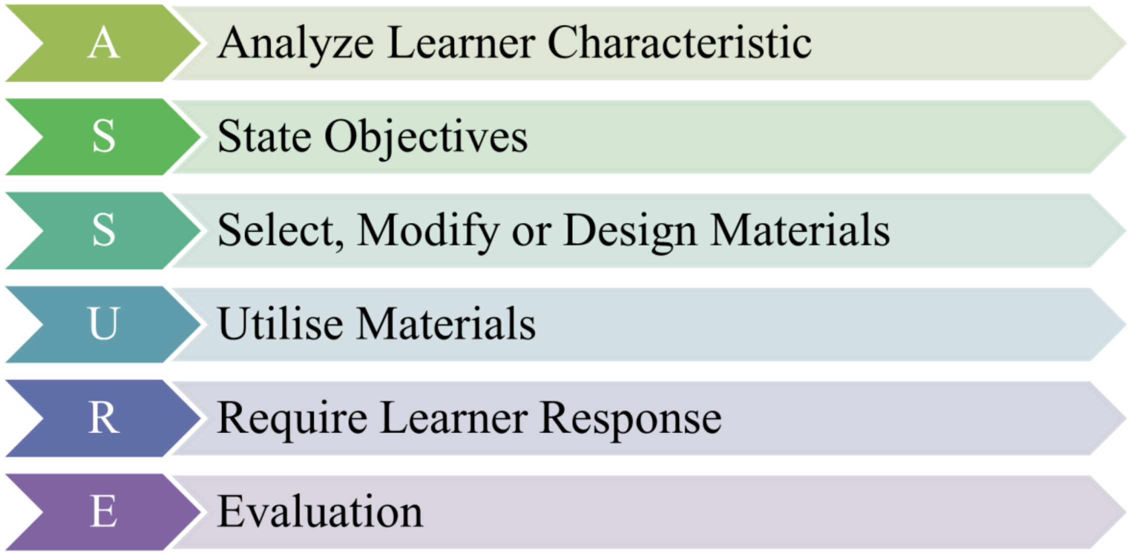


Figure 2.2: ASSURE Model (Heinich et al., 2002)

A – Analyse Learners

The teacher should begin by assessing the characteristics of her students as the first stage in the procedure. There should always be an emphasis on student characteristics that are linked to the targeted learning outcomes. The information acquired will assist you in making judgments about the remaining steps in the process. When educator knows what kind of learners they have, they may then choose specialised tactics and resources to help the students learn more effectively.

The followings should be recommended to include in educators' learner analysis:

- General characteristics of targeted learners such as gender, age, academic aptitude, interests, and others.
- Any prior knowledge and pre-existing skills
- Auditory, visual, and tactile learning modes are examples.

S – State Standards and Objectives

After analysing the learner's characteristics and preferences, the educator must then determine the instructional strategy's goals and conditions. This useful statement will epitomise what the instructional strategy can benefit the students. Next, the very same statement will concentrate on what the student can do eventually and gain understanding from the instructional strategy, to be more specific. Engineering students are an illustration of this. The learners will be able to list at least two data warehouses and multiple retrieval methods for finding building materials in a particular scenario.

The objectives can be used to evaluate the students' progress, possibly as part of the process to access student’s grade. The educator can also use them to communicate with students about the goals they will achieve during the class.

As an impression of a set of learning objectives with good quality is adherence to the ABCDs of clearly defined instructional objectives. The following are the details:

- Audience – The instructional objective is to target whom?
- Behavior – What was displayed in terms of performance or behavior?
- Conditions – Under what situation will the behavior or performance be observed?
- Degree – To what extent acquiring the knowledge or mastery level of skill?

S – Select instructional strategies

The another “s” from the model refers to select instructional strategies. In consideration of the fact that learning targets are set, the educators need to choose instructional strategies carefully that actually guide them to meet their targets.

To start with, the educator determines which teaching mode is most appropriate for her course. For instance, the weightage of instructional strategies to be attentive to the student and also how much to be weighted on the instructor? The preceding category includes instructional strategies like group projects, debate sessions and cooperative brainstorming. The latter category of instructional strategy example would be a demonstration, live lecture or multimedia presentation.

The instructional strategy around students is usually backed by common sense. The learning process will become much more engaging when class participation is welcomed. The student, not the educator, is climactically accountable for becoming proficient in the subject. Nevertheless, this can only happen when the teacher passes on an adequate level of critical material or coaches the techniques.

When the educator just guides the student toward discovering the correct answer for a problem rather than directly revealing the outcome will groom student independence so that the whole learning process is at its best. A good teacher is only a facilitator of the learning process; it should always be the student who is the one learning instead of the teacher.

After the educator has decided on her teaching strategy, she will need to find out which instructional strategies best support students over the span of their learning journey. Her responsibility covers all facets ranging from petty items such as writing tools to technologically advanced instruments such as projectors or touch screen display devices. It should be mindful that the lesson conductor herself is actually the most significant element of all in conveying instruction material. It is undeniable how useful are those intricate tools to transfer learning material, but a living individual who grasps knowledge beyond what a textbook can offer is most impactful.

U – Utilise instructional strategies

This step in the ASSURE process is to devise a strategy for implementing the instructional strategies chosen earlier on. The instructional designers must authenticate that their proposal, similarly to all other earlier stages aforementioned, will achieve the manifesto they have set to look upon.

To do this, it is very critical to embrace the "five P's" processes:

- Preview the instructional strategies

This implies that the instructional designers should think about how they are going to use those instructional strategies ahead of time. It's a good idea to do a trial or rehearsal before delivering it so that they can ensure that the whole lesson would execute steadily and without hiccups.

- Prepare the instructional strategies

Educators shall be collecting those teaching ingredients they need to deliver their instructional strategy. These instructional strategies must be in good functioning order. If educators are creating a PowerPoint presentation, for example, they are ought to develop the wording and visuals for each screen.

- Prepare the Environment

The medium in which education takes place, which is the environment, is compelled to be established with a handful of preliminary work. Even small actions such as making sure learners will have their own space for each of them is very vital. Additionally, in the circumstances where the teachers can have a say on how the environment should be set up, it is necessary for them to prevent students from being distracted from learning by any noise source, regardless it is loud or not.

- Prepare the Learners

To begin, teachers must clearly communicate the learning objectives to the students. This will assist students in creating a cognitive diagram showing them what are they going to study. The oncoming step is to inform the learners about how evaluation is being conducted on their learning results. Teachers must inform students of their tasks, how they will be graded, and whether or not tests will be administered. Teachers should also explain to the learners the advantages of mastering the content.

- Provide the Learning Environment

After that, it is time to put the learning into practice. This will be the stage where all previous preliminary work will bear fruit. Educators must always be aware that each of these preceding steps will lead to the whole learning process being managed. Henceforward, that educator will be considered successful in playing the facilitator role in the learning process.

R – Require Learner Participation

This step should have been included in the previous phases. It is of utmost importance that the educators draft a plan to get their students actively engaged with their learning content. As a means to that will require them to cater for the needs of not just the individual level but also the class as a group level itself.

Making the students participate actively in brainstorming or group discussions is one of the prime examples that educators can bring to the table. As a consequence, learners often will have to do their own preparation at home by getting a list of questions to be asked in class or takeaway to share not just to the teacher but also to their peers; this will cultivate interesting and robust learning. Another alternative to consider would be letting students display their confidence by leading the class or even sparking a thriving open discussion in a respectful manner such as a seminar.

Aside from that, educators need to figure out what are the factors that generally influence the students to take participation in the learning process. In what way are they going to acquire the knowledge and put the lesson learned in practice? This strategy must go beyond simply stating that they will listen and absorb the information. Perhaps the teacher can encourage a particular type of note-taking or other learning techniques to inspire the students.

E – Evaluate and Revise

The ASSURE procedure endmost phase is equally important as the others. Educators evaluate the impact of their teaching on learners with regard to their learning in current step. This comprises of an analysis of their teaching approach and last but not least the instructional strategies they employed. During this evaluation, the following questions are important to ask:

- Was the instructional strategy successful in achieving the learning objectives set out to achieve? How will we know if the students have met the objectives? Is our method of grading students in accordance with our learning goals?
- Is there any way to make this lesson better? How? What method will we use to evaluate the flaws in our presentation?
- Was our selection of instructional strategies wise? What criteria will we use to evaluate how effective these instructional strategies are?
- Could our instructional strategy possibly to have better performance?

The input from students should be the final step in the said evaluation. Was their whole experience positive? Do they believe they've accomplished both educators and their own personal goals? How will teachers know if students' performance was successful or not?

### 2.2.2 Instructional Strategies

Education facilitators utilise instructional strategies to aid students to learn to be independent and, at the same time, become strategic learners in the learning process. The strategy that steers learners to accomplish their work and meet learning objectives by allowing them to have a choice of approach they like to resort to and efficiently implement is the principle of instructional strategies (Mahmood, 2021).

Indication of how or what Instructional strategies can be used to or used as in the following:

- Helping them to gather their attention and thus focus and bring them inspiration
- Information can organise in a way that improves absorption and remember
- Perform an assessment on learning and become monitoring means

Conventionally, there would be five major divisions of instructional strategies, namely direct instruction, indirect instruction, interactive instruction, independent study and experiential learning.

*1) Direct Instruction*

Often known as the teacher-centered approach and it is a classical teaching method. In this case, the educator is the sole source of information for the students in the class. Lectures, didactic questions, power point presentations, and other similar methods are examples of this style.

*2) Indirect Instruction*

It is a kind of belief that learning is more meaningful when learners can seek out and discover information for themselves. A prime example of indirect instruction would be a debate, panels, brainstorming, collective investigation, and so on.

*3) Interactive Instruction*

Also known as Collaborative learning, in which students are usually in cluster grouping to engage in discussions, share their knowledge, express an opinion, explore questions together and other multi parties interact with each other.

*4) Independent Study*

This is a method in which the teacher instructs students to do something. The students will then undertake it on their own individually and consult their work with the teacher. It's also known as a teacher to student and student to teacher. Giving tasks such as essays, reports, projects, journals, and so on are examples of this strategy.

*5) Experiential Learning*

Lastly, this kind of strategy is to get students involved by hands-on experience with what they are learning. It appears that the real world is being experienced by students through the varied designs of activities created by the teacher.

Usually, the common society will decide that direct instruction is the least effective of the five types of instructional tactics because it does not interest students in learning, do not allow for contact between teacher and students, and does not allow students to communicate with one another. All activities revolve around the teacher. Those five types of instructional strategies, on the other hand, are incomparable. Each of them has a unique ability to improve the effectiveness of the class. It is undeniable that direct instruction can also be very successful when it comes to acquire information or a step-by-step skill. In addition, this stays true and effective too when the teacher is knowledgeable and skilled.

Thus, it all depends on what is being taught for a particular lesson and very much depends on how the teacher implements the instructional strategies, there are no definite best or worst as many factors, and aspects must be taken into account (Oakleaf & VanScoy, 2010). The followings are some examples for each kind of instructional strategy:

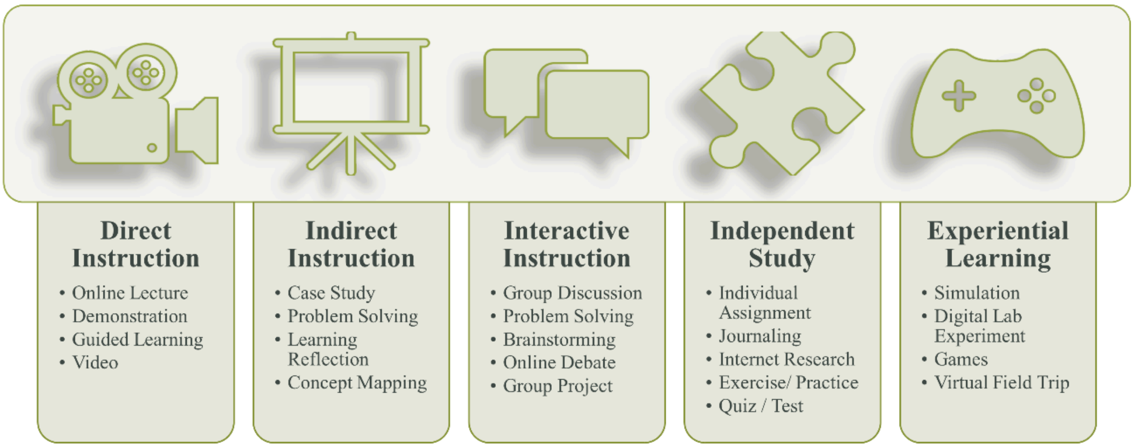


Figure 2.3: Instructional Strategies Kind

### 2.2.3 Online Learning

Unlike face-to-face learning, which allows teachers to go off-script, the entire process is pre-scripted in online learning, thus, designers must anticipate any potential hitches throughout the planning stages. In online learning, structure, content, and outcomes are more important.

The instructional design facilitates course creation while also ensuring that all learning objectives are accomplished. The benefits of the approach are realised by instructional designers during the production phase, while students feel them after the class is launched.

Okano et al. (2018) identified that among all learning content, the most popular learning method occupying a whopping 49% which is formal live classroom knowledge delivery led by an instructor or facilitator. However, with the rise of the internet and the advent of the COVID-19 pandemic, online learning has become much more popular than ever.

*1) Moving pre-existing content online*

For thousands of years, it is no secret that learning has always been a part of our lives, but much of the content we learn is predominantly grasped or apprehended in person.

When online learning originally became popular, one of the biggest issues was keeping students involved and participating in the lessons because so much of it was centred on reading and texts. With everything that technology now allows us to achieve, the first step is to find an authoring tool that suits our needs. Then we may begin customising the content from our in-person training to work online.

*2. Share information and generate interest*

Being able to communicate effectively requires practice and planning, especially in online learning. If educators are able to wrap up the knowledge students need in a compelling story, students will be more interested in the material and will be more receptive to the training and message. So, it is worth dedicating time to the story's script, and don't stop tweaking it until the message is teachers trying to communicate and meet the objectives they have set out.

*3. Encourage participation*

It is an absolute key that instructional designers research and understand their audience and how they learn best. Not all students are the same, so online courses must be able to cater to their various needs. Online learning doesn't give the teacher the freedom to go off script or tailor the material to their audience, so doing some research will go a long way. The online course must keep students participating with an engaging learning experience that utilises visual elements, compelling content, and comprehensive evaluation systems. Always bear in mind that if educators adapt to the needs of the students, the results will come.

*4. Ensure students retain what they learn*

By utilising fun, engaging elements such as video games, good instructional design not only gives students the material they need but also helps them retain it in the long run. Using multimedia at the right times lets students establish mental connections between the information and the actions it results in, which helps them better remember it.

*5. Motivate students*

If an online course is well-designed, students will be excited to be a part of it and will always find a way to make time for it. Good instructional design should answer the question: How do educators keep students motivated and wanting to learn more on the subject? Each e-learning course is different, so there's no one-size-fits-all formula teachers can apply to all classes. But, one thing is certain, be sure to design our courses with the target audience in mind.

In conclusion, the instructional design brings all the parts and pieces of successful online learning which is instructional strategies together so educators can oversee them in one place. This is undoubtedly the best way to maximise the benefits of online training in relation to face-to-face class sessions.

In addition to that, Nilson & Goodson (2018) pointed out that there are well above 7 out of 10 of targeted education institutions that grant degree accreditation offer online learning but the education courses usually struggle to integrate those best practices of pedagogy over online learning even though they are well equipped with technical resources. Consequently, this phenomenon will leave students having a hard time to keep up with pace and truly apprehend the education material, thus undermine their true potential of being a learner as these standards usually let go the best practice featured in learning and teaching as well as the cognitive science principles.

### 2.2.4 Personalised Learning

In psychology and cognitive science, the idea is that a person's past experiences will canonically contribute to shape their preferences in future. Ever since the discovery of classical conditioning, many researchers have discovered a variety of strategies to reinforce or reduce or convert the capacity of a person's preference for specific items using different experimental manipulations. Given a closer look at such research, it turns out that such research kind does not really mesh with all domains because the focus has not been on leading to preference being informed to model, instead it has been examining the associative nature of the animal.

VARK model is plausibly one of the most well-known theory developed to achieve personalised learning by identifying different learning styles of learners. Based on Fleming & Baume (2006), everyone can be categorised into at least one of the four arch types of learners, namely visual, auditory, reading/writing and kinesthetic. In this research project, the VARK model is employed as the main philosophy of the questionnaire to designate learners accordingly so that it can serve as a source of personalisation variables for classification models. The following figure demonstrates some examples for each of the learning styles in the VARK model:

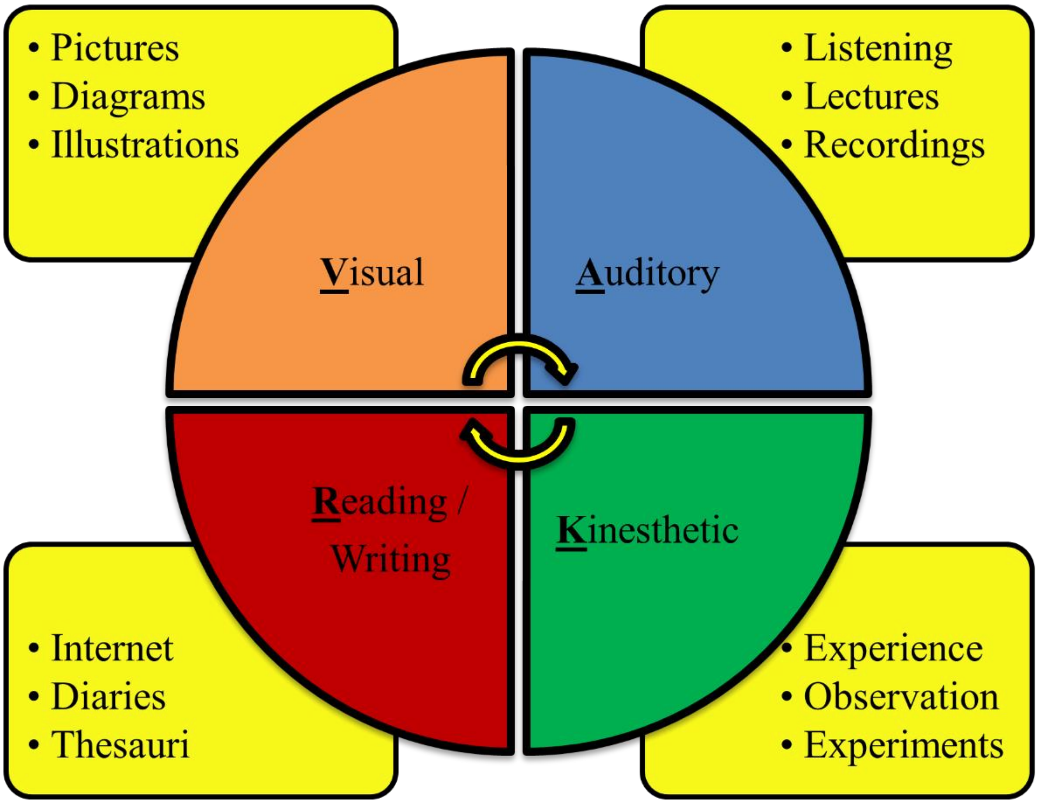


Figure 2.4: VARK Model and Examples

Undoubtedly, it is widely accepted that we humans form our preferences via experience when we are in a situation where both data and beliefs are absent, thus, personalisation is an essential input to model (Srivastava & Schrater, 2015). In education, personalisation exists through the study of learners' preference on how the learning process takes place. Thus, personalisation can also be expressed as an educational tactic that aims to customise the learning process based on the learner's needs, strengths and interests. As pointed out by Kaya & Altun (2011), it is possible to provide students with a personalised learning experience that makes use of various systems and environments. They lament that most of the models available are either too bland or too sophisticated to apply especially when it comes to online learning.

From the pool of researcher's literature, one can see that majority of them proposed whole-person personalisation, which advocates the combination of both psychological resources and cognitive personalisation. Personalisation allows the system to actually reveal the characteristics of the learner (Ghallabi et al., 2013). With the aim of achieving learning content and activities personalisation, the selection or customisation of learning content and activities is very critical. Personalisation is commonly used to refer to a learner's learning behavior, emotional state, and his knowledge level. As discussed in the previous section, there are numerous instructional design models that can be used for personalisation. In the opinion of Gamalel-din (2012), personalisation can take many forms, such as personalised learning environments, personalised learning content, and personalised interfaces.

## 2.3 Review of Classification Algorithms

A classification algorithm, also known as a classifier, can be defined as an algorithm that will take a collective of features into consideration and then proceed with categorising objects before utilising them to predict the object class. Generally, in machine learning, there are two types called supervised and non-supervised machine learning (Ayodele, 2010). In supervised machine learning, the machine uses a human expert to classify the object and produce a sample of objects with identified classes, commonly known as labelled data. We call this group of identified objects as the training set reason being it learns how to classify objects based on classification programs. Unlike unsupervised machine learning, the machines instead directly work with the data, and the main difference here is none of the training data features prior to assigned classes, better known as unlabelled data. In this research project context, classification is all about a supervised machine learning technique in which the machine will learn those labelled data variables using training data. To put it another way, classification refers to model that predicts outcome by making of labelled data to draw eventual conclusion. Further section reviewed the past papers which applied classification algorithms that are used in this study based on whether they are supervised machine learning model that fit our research problem or not.

### 2.3.1 Gradient Boosting

Back in the day decision tree was a very popular machine learning algorithm, and it still continued to enjoy gargantuan popularity because decision trees are easy to interpret. It works seamlessly to visualise those data patterns that are nonlinear. In addition to that, the decision tree is economically cheap to run for a small dataset. Inversely, the performance is not great when it is used to deal with a messy or big data set.

Then comes to the idea of boosting, and many researchers are starting to study how to convert those relatively weak decision trees into an efficient algorithm. Gradient Boosting (GB) is one of the most prominent frameworks developed by Friedman (2002) that optimises a particular loss function by making use of weak learners such as decision trees to do predictions in an ensembled manner additively. Madeh & El-Diraby (2021) published a paper pointing out GB usually has better performance when pitched against Random Forest.

$$F_0(x) = \arg\min_{\rho} \sum_{i=1}^{N} L(y_i, \rho)$$

For m = 1 to M do:

$$\tilde{y}_i = -[\frac{\partial L(y_i F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}, i = 1, N$$

$$a_m = \arg\min_{a,\beta} \sum_{i=1}^{N} [\tilde{y}_i - \beta h(x_i : a)]^2$$

$$\rho_m = \arg\min_{a,\beta} \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + \rho h(x_i; a_m))$$

$$F_m(x) = F_{m-1}(x) + \rho_m h(x; a_m)$$

end For

end Algorithm

Figure 2.5: Gradient Boosting Algorithm

It is worth noting that GB's training process customarily requires a longer time because it updates a large number of parameters in its lengthy process (Cai et al., 2021).

### 2.3.2 Naïve Bayes

Being one of the staple examples of supervised learning algorithms, Naïve Bayes (NB) makes an assumption of independence from each other among every feature, meaning that they all have equal importance, in other words. Catal and Nangir (2017) attest that NB algorithm could still perform well even though the features are not independent. The whole foundation of NB is centred on probabilistic theory, and it is a relatively simple and easy to understand algorithm.

$$P(C_i|x_1, x_2 \dots, x_n) = \frac{P(x_1, x_2 \dots, x_n|C_i).P(C_i)}{P(x_1, x_2 \dots, x_n)} \; for\; 1 < i < k$$

Figure 2.6: Naïve Bayes Theorem

In spite of being an easily understood algorithm, NB has several notable down sides and one of them being ineffective when dealing with multinomial models with high dimensional space (Wong & Tsai, 2021).

### 2.3.3 Support Vector Machine

What full name SVM does an attempt to search for the hyperplane with the greatest margin that potentially delivers the maximum separation among all class members and on top of that, those samples that are nearest or closest to the said hyperplane are dubbed as support vectors (Catal & Nagir, 2017).

As attested by Maroco et al. (2011) that SVM, alongside another widely used algorithm used for classification problems. Neural Networks, their performances are enormously dependent to features selected and also the value of parameters tuned.

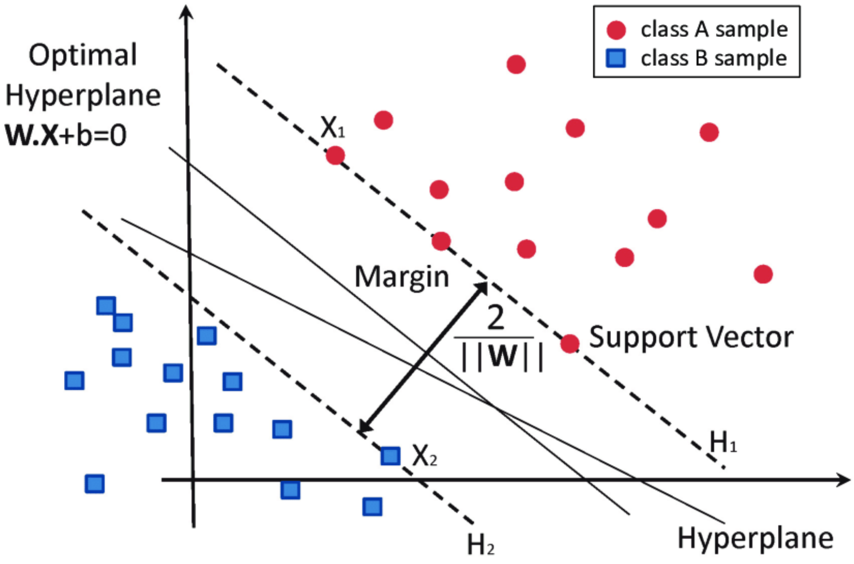


Figure 2.7: Support Vector Machine with Hyperplane (García-Gonzalo et al., 2016)

## 2.4 Literature Map

It goes without saying that how critical it is to plan out what topics that will be covered as well as establish a connection between them. Researchers desire the literature review to feature an overall logical structure that plays out a well-defined and concise theme pivoting related ideas to be explored and developed around. A literature map is definitely worth the time to prepare, considering the valuable productivity one could gain from it by having a graphical presentation to various themes and how they are connected to each other.

Coined by Creswell (2014), a literature map introduces information in diagram form using in two-dimensional style in which linkage between information can be tied together by adopting lines or perhaps with the aid of text but not necessarily a must. The benefits of producing a literature map are as follow:

- Gives us a better visual to deduce key component and literature research findings
- Systematically arrange of our ideas and thoughts
- Appealing mapping network to group findings alike

The very same literature map also serves as our guidelines for literature review.

The following figure is the literature map for this research project:

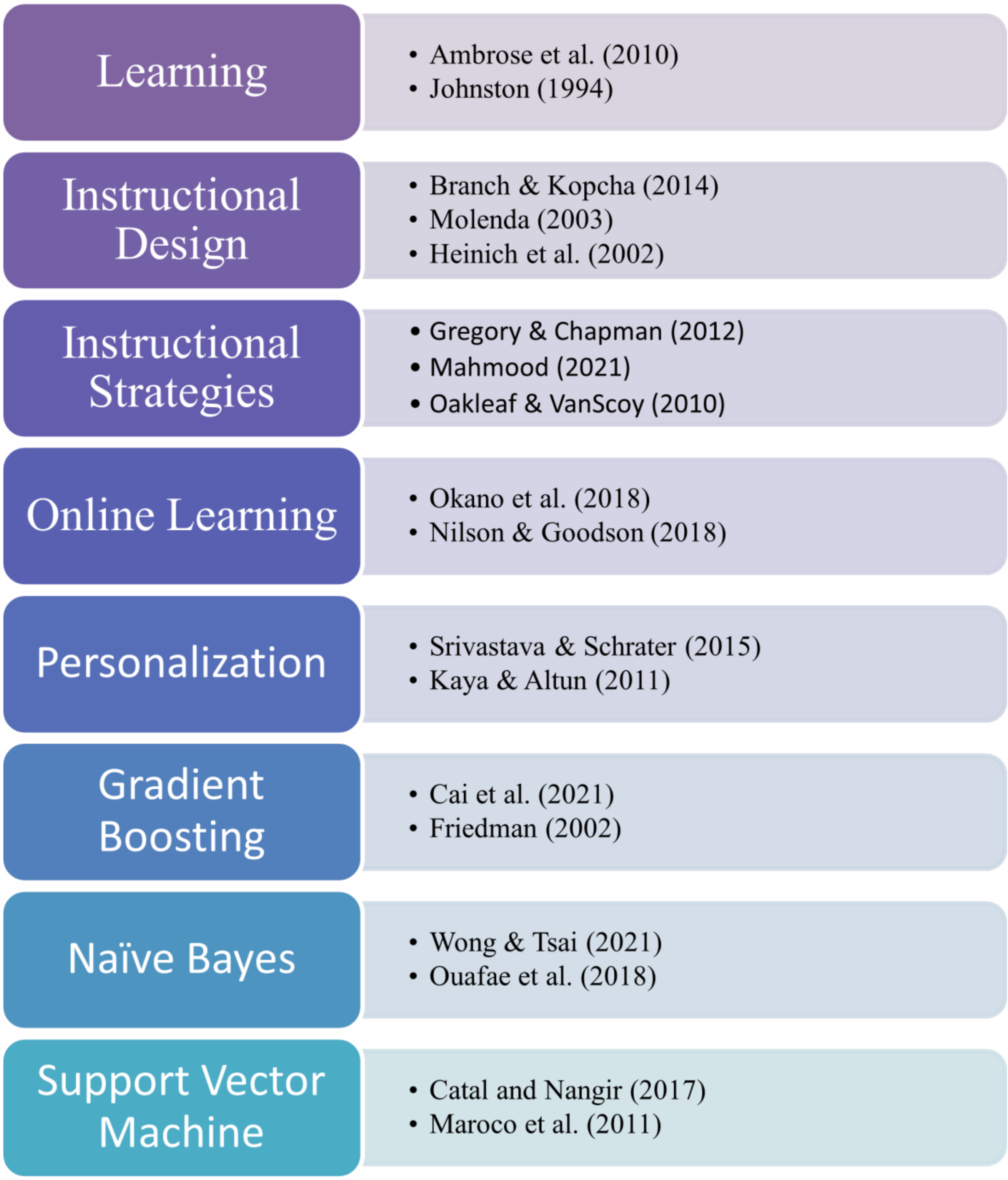


Figure 2.8: Literature map

## 2.5 Previous Work in Classification of Online Instructional Strategies

There is no lack of use of machine learning on classification problem as one can witness many studies has been carried out throughout the years. However, there are not many researchers who proposed classification of online instructional strategies as most of the studies focused on instructional design instead.

Meanwhile, a paper written by Eristi & Ardeniz (2012) drafted a four-phase process to act as a scale to diagnose instructional strategies. This study does not explore the concept of personalisation and the chief motivation is diagnosing instructional strategies with a scaling approach rather than classifying them.

Despite that, Li et al. (2021) published a paper that encompasses personalisation into federated learning which is nothing to do with education, yet this shows that researchers are starting to take personalisation into machine learning problems recently.

Lastly, Yavgildina & Mishina (2015) discussed about personalisation for content to be delivered to students who are specialised in artistic pedagogy only. The personalisation aspect is limited to instructional objects rather than applying on the classification of instructional strategies.

## 2.6 Summary

In this section, the topics were explicitly reviewed. Based on the literature review, problem statements and research problems were derived. The findings of this research suggest the importance of personalisation in learning process. Learning style plays an impactful role, especially for learner or student. While students acknowledge their own learning style, it is very important that whether they possess the ability to incorporate it into their learning process resulting them able to gain experience from that. Hence, the quality of their learning process will be way better and effective, at the same time, experience enrichment would be much more fulfilling and satisfying. Furthermore, students are accountable to their own education, so those who put in effort to understand their learning preference will reap the benefit as compared to those who does not commit.

Indirectly, learners' confidence will be boosted and more encouraging as a positive side effect as well. As an educator who integrates personalisation into instructional strategy during designing stage can provide students with a better learning experience. Besides that, proper handling with care of instructional mismatch will cut down students' dropout as it is worth acknowledging the soaring of dropout rates that was even more prominent since the advent of COVID-19 pandemic. The next chapter discusses the methodology taken during this research project.

# CHAPTER 3: METHODOLOGY

## 3.1 Introduction

In this chapter, we succinctly discuss how this study is to be tackled to achieve those objectives as well as solve those problems detailed in the section of problem statement. The methodology of online instructional strategy classification is illustrated in Figure 3.1 as follows, where the entire research design of this research project is primarily divided into five main processes:

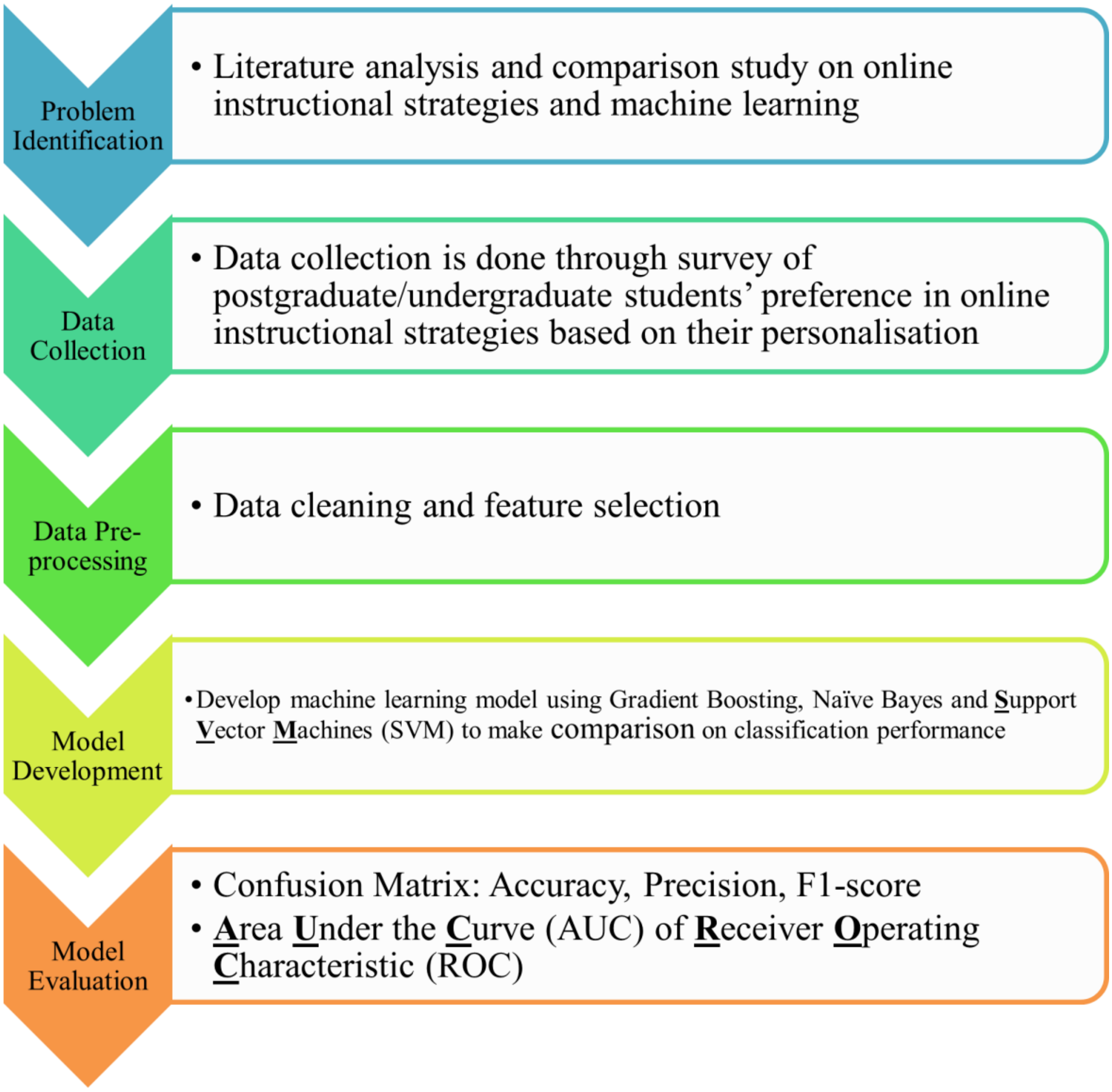


Figure 3.1: Main Processes of Research

## 3.2 Problem Identification

In respect to problem identification, it mainly involves a literature review and preliminary study. The research problem is identified by reviewing previous works through literature analysis related to instructional strategy, instructional design, the concept of online learning, personalisation, machine learning and information extracted by comparing the only few previous works published.

## 3.3 Data Collection

A questionnaire survey was conducted and distributed to primarily postgraduate and undergraduate students as the unit of analysis. For this research project, convenience sampling was applied, resulting in the cumulative number of 985 respondents. Convenience sampling refers to a sampling technique in which the type is non-probability whereby the sampling process is performed on a crowd of individuals that are contactable and reachable. A paper by Etikan & Alkassim (2016) explained that in convenience sampling, a sample is selected from elements of a population who are willing to be studied and are easily accessible. The respondents provided a reasonable representative profile, and responses were obtained from several genders, faculty, level of study, the field of study and many others.

This questionnaire consists of a four-page self-administered questionnaire. It comprises the following four parts:

- Part 1 (Personal Details) – Consists of demographics profile consists of e-mail and followed by basic details such as sex, country origin and household income. Then, Education level and other related information such as what do they study and which faculty are they belong to as well as which learning institution was being surveyed before moving the hindmost part dealing with their preference over learning mode, a communication platform, what they frequently used social media as well as their opinion concerning online learning difficulties (twelve questions).

- Part 2 (Learning Object) – Indicating students' preference to use the listed learning objects and online instructional strategies in their study (two questions).
- Part 3 (Learning Style Awareness) – Indicating students' awareness and understanding of their learning style (three questions).
- Part 4 (Learning Style Test) – Testing students about learning style based on VARK model to provide the basis for personalisation (thirty questions).

Note: Incentives were not provided to respondents for completing the questionnaire.

## 3.4 Data Pre-processing

Regarding the data cleaning stage, it is carried out after the content from the respondent is extracted. Data cleaning constructs the content for the feature extraction process later by removing incorrect, incomplete or outliers which result in incorrect feature identification. Acknowledging the substance of data pre-processing in producing a magnificent quality dataset, this research project has undertaken extensive data cleaning.

### 3.4.1 Data Cleaning

Data cleaning was done on this research project dataset by handling those missing values and ambiguous results. Due to the nature of survey data, data cleaning is crucial in identifying and removing responses which either out of target respondents or not answer questions thoughtfully. The following are the example of data cleaning done on the dataset.

- Removal of those respondents who do not meet the target criteria such as non-postgraduate or non-undergraduate students.

- Removal of those respondents who only answer a portion of questions or leave blank or answer inappropriately.
- Reassign as NA for the ambiguous response, doubtful results, no result especially in the open-ended question.
- Amend those response with notation or symbol or special character such as "?", "@", and "/".
- Z-score standardisation was used to rescale the dataset for the purpose of improving the model's performance.

### 3.4.2 Feature Selection

According to Sun et al. (2004), there will be many features that could be deemed as redundant or end up being irrelevant if one does not employ any sort of strategy to select features to do classification task.

Initially, the dataset originally consisted of 103 variables. The variables that were not contributing to the prediction, such as timestamp and identification, were being removed accordingly. With the bag of words model, there will be massive features generated in the dataset. However, since this research project is doing the classification of online instructional strategies, it only selects those related features and drop those columns did not need such as learning objects. The central motivation of feature selection is to achieve good accuracy results in the machine learning model by using the least possible features. A lot of features might not be beneficial for machine learning as this may cause over-fitting to the models.

In this study, the feature selection method is used by computing mutual information or be called information-gain. Information gain measures how much information a feature gave in order to predict the class, and it is measured by the reduction of entropy. Entropy can be described as the measure of uncertainty. Henceforth, the more reduction on the entropy (uncertainty), the more information provided from the feature. Eventually, the feature selection method will only choose those features with high information gain to build the machine learning models. In the end, 40 target variables and 30 feature variables were being selected for this research project.

## 3.5 Model Development

This research project mainly based on the classification using supervised machine learning, whereupon basically categorises a group of data into classes to predict the class of the unlabelled data. Training data consisting of labelled data was used to train the classifier to understand how to input variables that are related to the target. The experiment divides each dataset into two sets which is the training dataset and the testing dataset at the scale or percentage of 80% and 20%, respectively. Cross-validation is performed using the K-fold cross-validation technique. For this purpose of this study, we will perform ten folds cross-validation to evaluate our algorithm performance. The idea of ten folds cross-validation technique is to split the dataset into ten groups after shuffling the dataset randomly, then proceed to take each of the ten groups out on their own to serve as test data set while the rest will serve as train data set and analyse the performance of algorithm after fitting the said ten newly grouped dataset. Then the ultimate result of a single estimation is derived from the computation of all performance matrixes average. In a nutshell, there will be three supervised machine learning models being used to perform classification of online instructional strategies for this research project, namely Gradient Boosting (GB), Naïve Bayes (NB) and Support Vector Machine (SVM) as they are suitable for classification problems and relatively easier to implement.

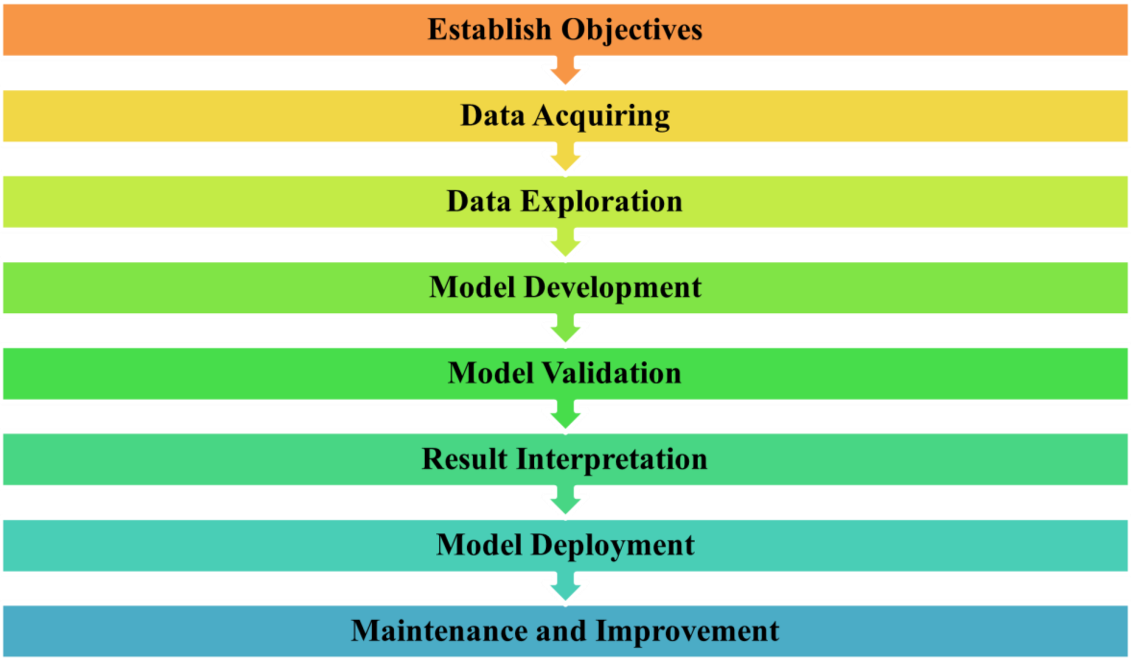


Figure 3.2: Model Development Process Flow

### 3.5.1 Gradient Boosting

As discussed in section 2.3.1, GB is a prediction algorithm that combines multiple weak learners that is usually one level attribute splitting decision tree called a decision stump to mitigate the loss function by taking advantage of Gini coefficient before making the prediction. Due to its ensemble nature, GB being costly is inevitable. We will be installing GradientBoostingClassifier from the ensemble module of scikit-learn Python library as our GB model for this research project:

```
import sklearn
from sklearn.ensemble import GradientBoostingClassifier
```

Figure 3.3: Python code to import GB

### 3.5.2 Naïve Bayes

It is no secret that why Naïve Bayes is so frequently being deployed in many machine learning models for classification algorithm because it is simple to use and very cost effective. It utilised the Bayes theorem to determine a class of the object belongs by computing the probability. In this research project, Naïve Bayes will calculate the probability of the learning object belongs to which personalisation. It will then classify the learning object into the class with the highest probability. The Python code for importing NB model from scikit-learn library is as follows:

```
import sklearn
from sklearn.naive_bayes import GaussianNB
```

Figure 3.4: Python code to import NB

### 3.5.3 Support Vector Machine

SVM is a non-probabilistic classifier that can be used in classification. How SVM operates is breaking up data by using line boundary dubbed as hyperplane that is calculated by only a handful smaller set of the data referred as support vector. The boundary is called a hyperplane and will be optimised for the purpose of getting the maximum separation (margin) between the two classes. SVM then uses that hyperplane in order to predict the class of a new data object once conferred with its feature vector. The following figure shows the Python to import SVM from scikit-learn library.

```
import sklearn
from sklearn.svm import SVM
```

Figure 3.5: Python code to import SVM

## 3.6 Model Evaluation

Once we get all the models trained, every single model will then be being tested with the test data. The performance of the model will be judged and measured from the accuracy, precision and f-1 of the prediction, which can be computed by formula using component from the confusion matrix below.

Table 3.1: Confusion Matrix or contingency table

| | Actual Positive | Actual Negative |
|---|---|---|
| Predicted Positive | True Positive | False Positive |
| Predicted Negative | False Negative | True Negative |

The first evaluation metric will be looking at accuracy which allows us to measure how accurate our model predicts. Those prediction outcomes which are true (both true positive and true negative) represent the correctly predicts will compare against the whole data set to give us an idea how accurate the model is as we can see from the below formula:

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + False\ Negative + True\ Negative}$$

Coming up next is Precision. This metric will exhibit the percentage of correct prediction over the relevant label. In layman terms, it substantiates how good the model at predicting the relevant class that we are interested, disregarding those outcomes that we do not desire. The following is the formula for precision:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

The next metric for evaluation is F-score and this research project will use the standard F1 instead of other adjusted F-scores such as F0.5 or F2 in this research project. It is an average that represent a ratio of precision reciprocal with recall, the following formula derived based on mathematical concept of harmonic mean:

$$F1 = 2 \times \frac{Precision\ \times Recall}{Precision + Recall}$$

As seen from the formula of F1 score, calculation of recall plays a part in it. Recall is also sensitivity in some literature, basically it indicates what is the true positive rate. For this research propose, recall will not be used as evaluation metric but the formula is needed for F1 formula understanding:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

### 3.7 Summary

To conclude, this chapter addressed the model development process that starts from the introduction all the way to the framework and experiment evaluation. On top of that, this chapter also underlined some related as well as important issue to tackle the improvement of model performance. The next chapter presents in detailed discussion on the study and evaluation of the model performances. It is crucial to acknowledge that a deeper understanding of the results will assist in verifying the suitability of the framework in facilitating the classification of online learning strategy.

# CHAPTER 4: FINDINGS AND DISCUSSIONS

## 4.1 Introduction

There will be two major parts making up this chapter, which are findings and discussion. The first section illustrates findings derived from model development for the classification of learning strategies using machine learning algorithms. Next, will be describing the data before moving on to elaborate on data pre-processing and also evaluate the model performance in the discussion section.

## 4.2 Findings

### 4.2.1 Part 1: Data Exploratory Analysis

A questionnaire has been distributed accordingly to undergraduate and postgraduate students as per Section 3.3, and the following table depicts the respondents' distribution based on their demographic profile.

Table 4.1: Demographic Profile

| **Demographic Profile (n = 985, Malaysian = 969)** | | **Responses (N)** | **Responses (%)** |
|---|---|---|---|
| **Gender** | Female | 661 | 67.1 |
| | Male | 324 | 32.9 |
| **Household Income** | Less than RM 4,849 | 587 | 59.6 |
| | RM 4,850 – RM 10,959 | 281 | 28.5 |
| | More than RM 10,960 | 89 | 9.0 |
| **Level of Study** | Certificate/Diploma | 134 | 13.6 |
| | Undergraduate | 834 | 84.7 |
| | Postgraduate | 14 | 1.4 |
| | PhD | 3 | 0.3 |
| **Field of Study** | Computer Science / Information Tech | 156 | 15.8 |
| | Architecture and Building | 107 | 10.9 |
| | Engineering | 92 | 9.3 |
| | Economy | 68 | 6.9 |

| | | | |
|---|---|---|---|
| | Social Science | 63 | 6.4 |
| | Education | 62 | 6.3 |
| | Accounting and Finance | 62 | 6.3 |
| | Medical | 52 | 5.3 |
| | Business Administration | 50 | 5.1 |
| | Linguistic | 48 | 4.9 |
| | Sports | 44 | 4.5 |
| | Arts | 41 | 4.2 |
| | Religious Studies | 39 | 4.0 |
| | Others | 101 | 10.3 |
| **Institution** | UM | 555 | 56.3 |
| | UITM | 154 | 15.6 |
| | KKTM | 91 | 9.2 |
| | Others | 185 | 18.8 |

Thereafter, a series of graph will be used to present the various preference collected from the said survey in which respondent can answer more than one choice:

- Preferred Learning Mode

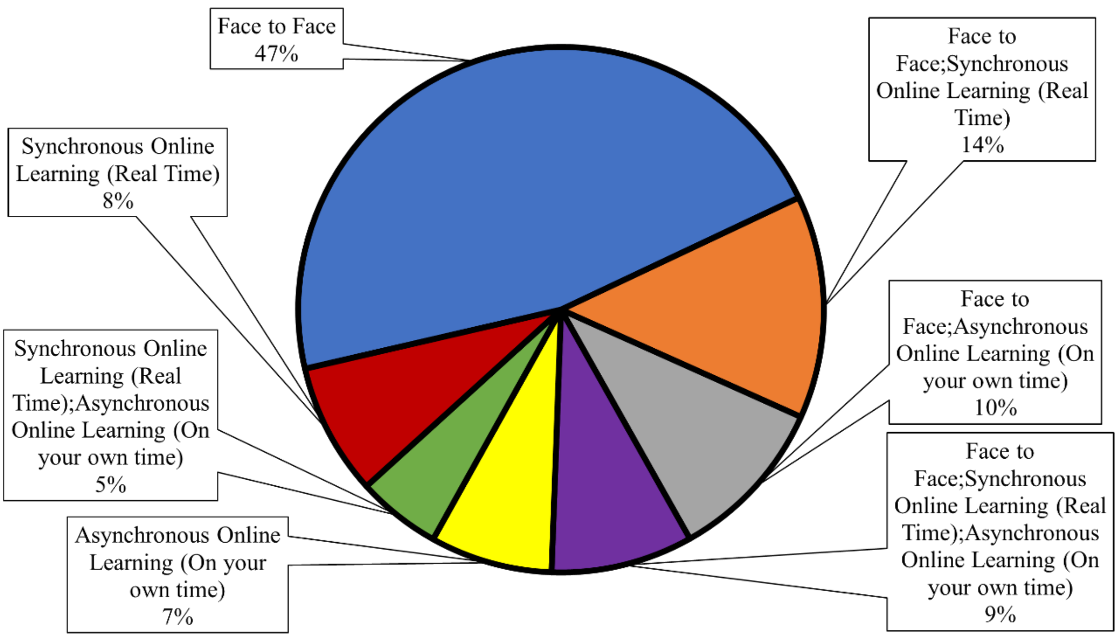


Figure 4.1: Preferred Learning Mode

As depicted by the pie chart that face-to-face is still the most preferred learning mode that alone accounted for 79.2% in all of them at least having it as one of the choices selected.

- Preferred Social Media Platform

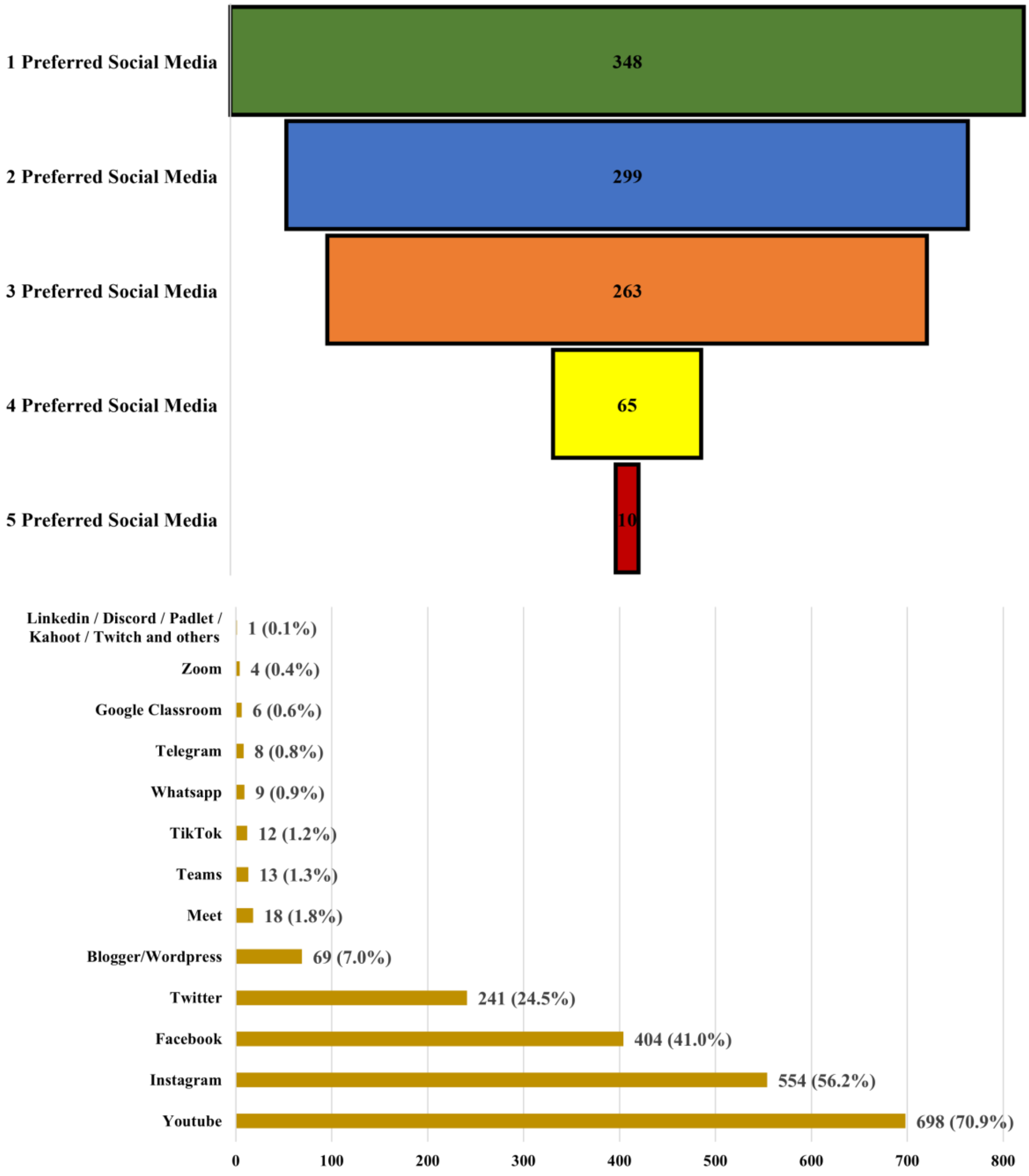


Figure 4.2: Preferred Social Media Platform

Everyone is living in modern era that which social media has already become an integral part of human lives. From the figure above, one can reckon that having more than three social media platforms can be very time-consuming to manage them all actively and there is the possibility that growing number of social media platforms can perform more or less the same task the others can offer, thus having too many social media platforms can be redundant too.

- Preferred Communication Platform

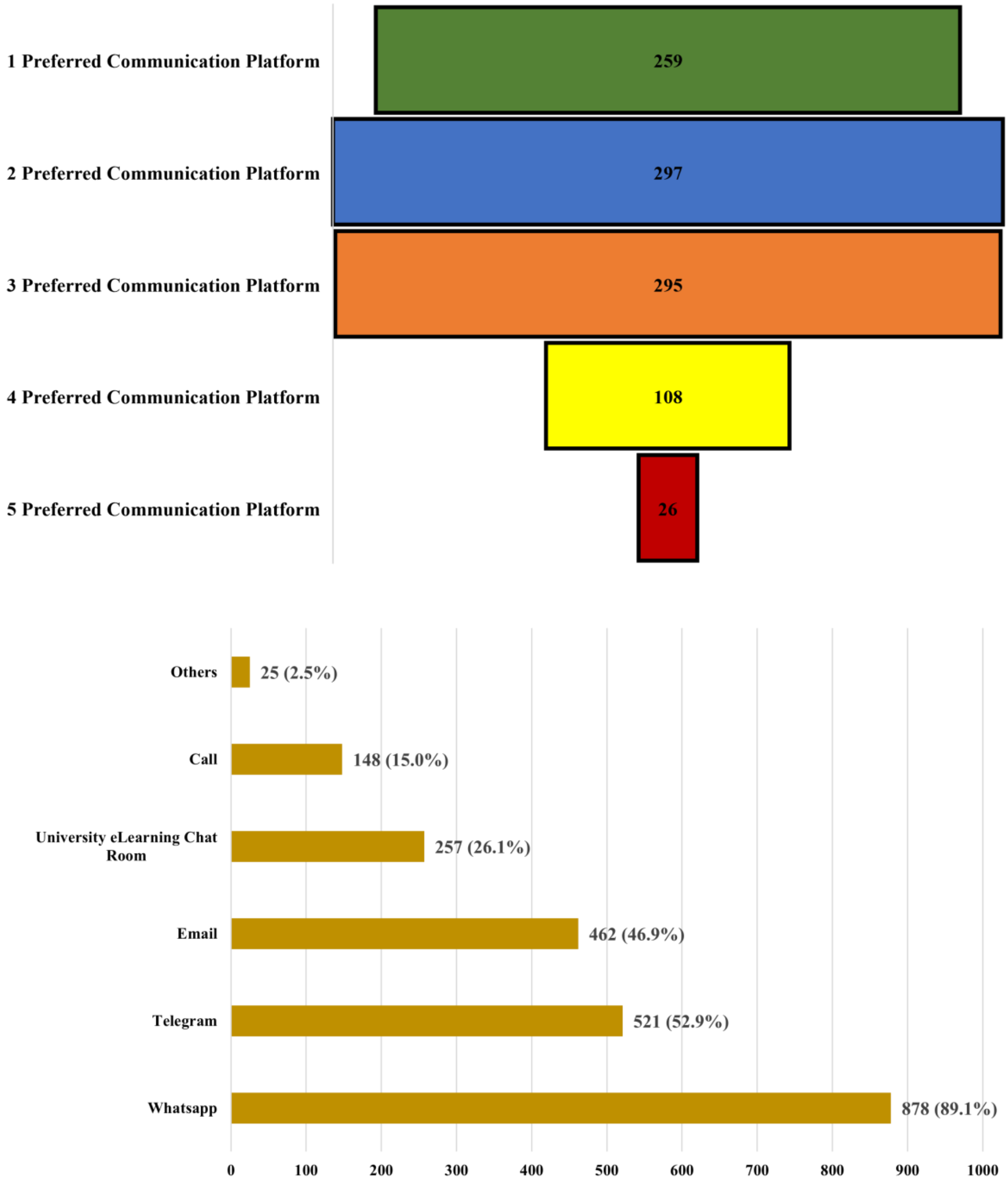


Figure 4.3: Preferred Communication Platform

As for this question, it is to address students' preferences in communication during their learning process. From the figure, it is obvious to identify Whatsapp is overwhelmingly popular and the gap between the next popular is so big, showcasing that Whatsapp is almost equivalent to what English is in the linguistic world as time is the utmost factor when it comes to quick communication, so having the most widely used communication platform will cut down the time spent and save the hassle to install alternatives on one's device.

- Difficulties in Online Learning


Figure 4.4: Difficulties in Online Learning

This question demonstrates difficulties faced by students in online learning. There was a free text cell for respondent to type in their opinion on top of the existing 11 selectable options available. The above word cloud illustrates those common difficulties faced by student in online learning weighted by the size of the word font.

- Instructional Strategy

This study reveals preference of learners over instructional strategy implemented by educators. There were five levels of scale selection for each of these questions ranging from “Not At All” until “Very Much”.

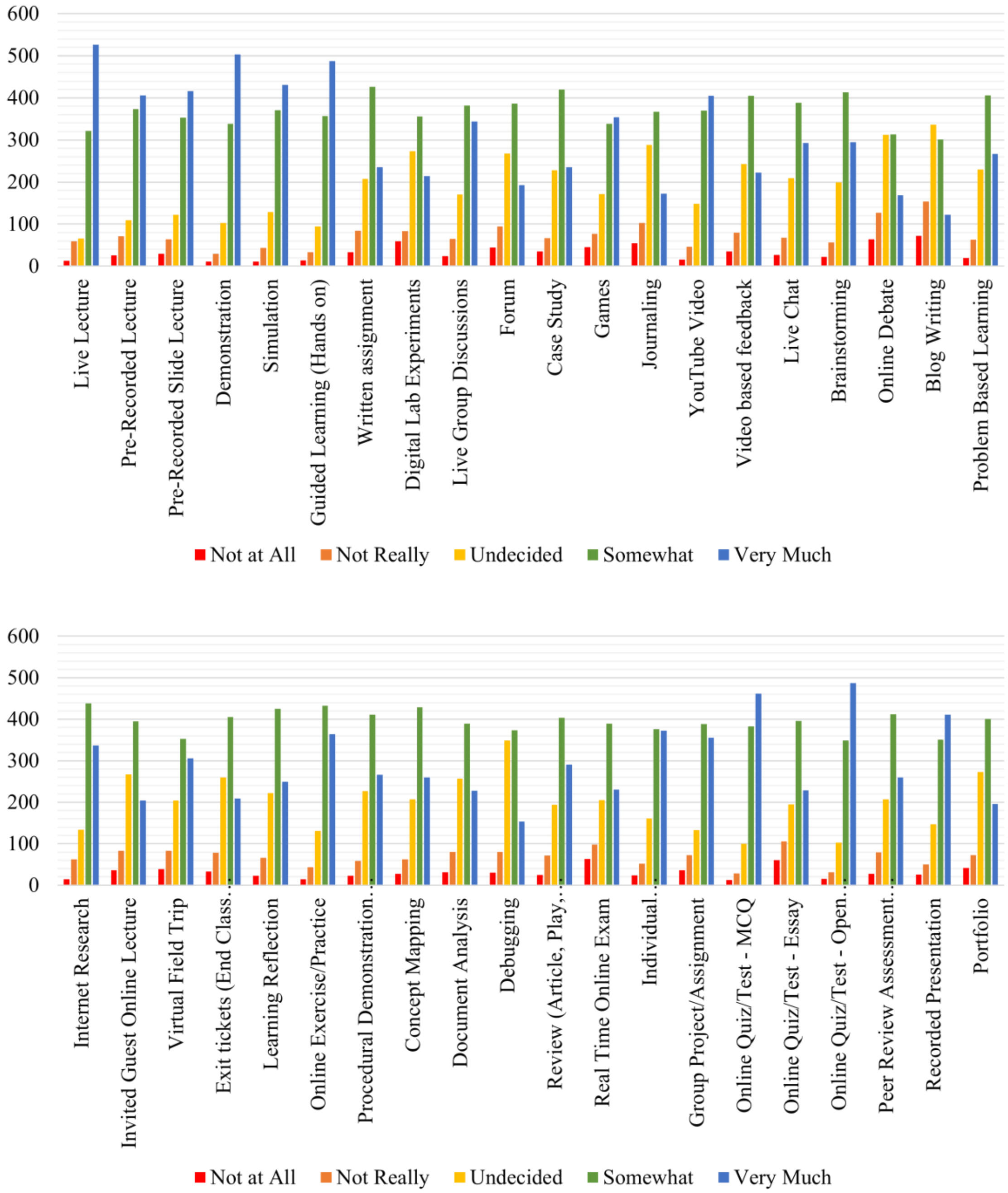


Figure 4.5: Instructional Strategy

- Learning Style Awareness

Regarding those awareness questions, respondents generally think that learning style is important to improve their learning ability. Some of them do know their preference even though they are not familiar with term learning style before survey.

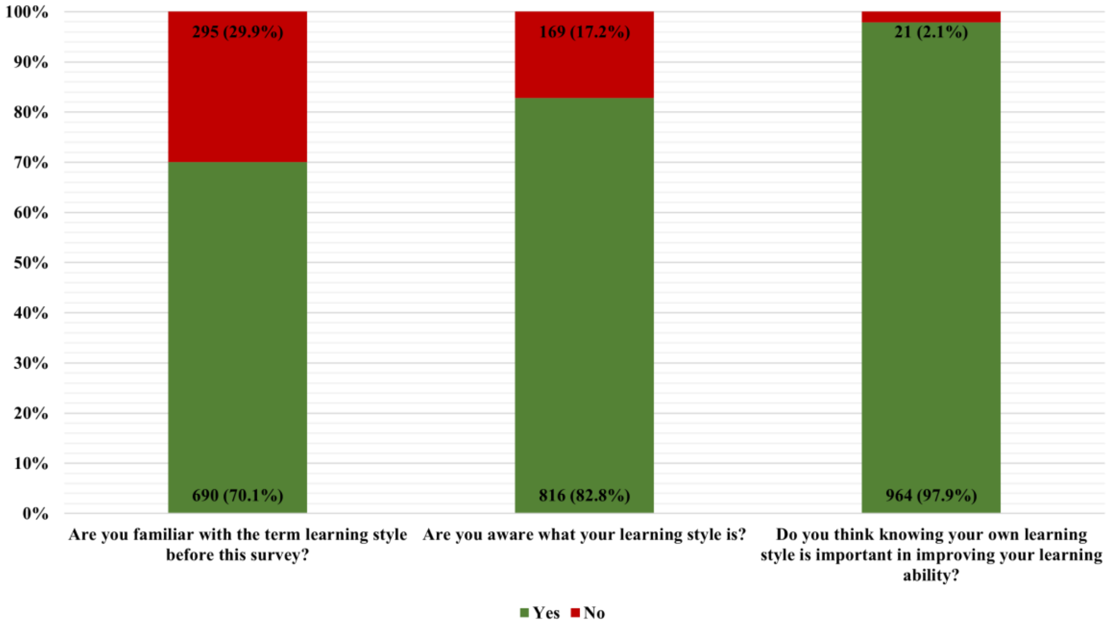


Figure 4.6: Learning Style Awareness

### 4.2.2 Part 2: Data Modelling

This section comprises of two parts, first will deal with data partition, which is to split our dataset into train data and test data, respectively. After that, will then proceed on to use the preceding one to develop and train all three of our models. Then at the end of this section, the performance of all three models is evaluated and a comparison will be made to pinpoint declare the best model.

#### 4.2.2.1 Data Partition

With regard to the data partition, it will be splitting the dataset into four-fifth of it as training data and the remaining one-fifth as testing data as per what being discussed in Chapter 3, section 3.5. Original sequence or index of the original dataset were not being followed instead; randomly selection will take place to determine what goes into training data and those remainders would be our testing data before proceeding on to perform K-fold cross validation. The purpose of performing data partition also known as splitting of data is to have training data to train the predictive model and oppositely the testing data to test the model accuracy.

The following depicts the data partition and the outcome amount after the python code:

## Train Test Split

In [15]:

```
xtrain, xtest, ytrain, ytest = train_test_split(TrainDF,TargetDF,test_size = 0.2) #splittin
ytest = ytest.values.tolist()
```

In [16]:

```
np.shape(xtrain),np.shape(xtest),np.shape(ytrain),np.shape(ytest)
```

Out[16]:

```
((788, 30), (197, 30), (788, 40), (197, 40))
```

Figure 4.7: Data Partition with Python

#### 4.2.2.2 Model Building

As discussed earlier in the previous section, there will be three models being developed to predict instructional strategy using Python programming language. After data partition takes place, will start the next process to train each of the models for each target variable which is an instructional strategy, before stacking them in an ensemble manner.

## Model Training

In [17]:

```
def GetMetrics(ytest,Pred,model,xtest):
    xtest = xtest.values.tolist()
    ListF1s = []
    ListPrecision = []
    ListAccuracies = []
    for index in range(0,len(Pred)):
        ListF1s.append(f1_score(ytest[index],Pred[index],average='micro'))
        ListPrecision.append(precision_score(ytest[index],Pred[index],average='micro'))
        ListAccuracies.append(accuracy_score(ytest[index],Pred[index]))
    return ListF1s,ListPrecision,ListAccuracies
```

Figure 4.8: Python code to train models and storing of each iteration result

Same logic applies to all three models from GB to NB to SVM and the following figure shows the Python code for prediction process:

## Gradient Boosting Classifier

In [19]:

```
GB = GradientBoostingClassifier(learning_rate=0.1)
Multi_Label = MultiOutputClassifier(GB)
model = Multi_Label.fit(xtrain, ytrain)
Pred = model.predict(xtest)
```

## Naive Bayes

In [23]:

```
GNB = GaussianNB()
Multi_Label = MultiOutputClassifier(GNB)
model = Multi_Label.fit(xtrain, ytrain)
Pred = model.predict(xtest)
```

## Support Vector Machine

In [27]:

```
ModelSVC = SVC(decision_function_shape='ovo')
'''
Reason For Using OVO
In this method, every single class will be paired one by one with other class.
At the end of the classification training,each classification is given one vote for the win
The highest votes will determine which class of each label the test dataset belongs to.
'''
Multi_Label = MultiOutputClassifier(ModelSVC)
model = Multi_Label.fit(xtrain, ytrain)
Pred = model.predict(xtest)
```

Figure 4.9: Python Code for Model Prediction

For this research project, SVM decision_function_shape parameter will be set using a one-vs-one decision function which is to train a classifier for each of the class pairing combinations instead of using a one-vs-rest that will train one classifier for every class fit against the rest of the classes.

Once all three models are done, a data frame table will retrieve those results and present them in tabulate form to the comparison of model evaluation metrics using accuracy, precision and F1 score as per the following figure:

In [31]:

```
dff = pd.DataFrame({'Model':["Gradient Boosting","Naïve Bayes","Support Vector Machine"] ,
dff
```

Out[31]:

| | Model | Accuracy | Precision | F1 |
|---|---|---|---|---|
| 0 | Gradient Boosting | 70.697970 | 70.697970 | 70.697970 |
| 1 | Naïve Bayes | 68.946701 | 68.946701 | 68.946701 |
| 2 | Support Vector Machine | 72.043147 | 72.043147 | 72.043147 |

Figure 4.10: Accuracy, Precision and F1 Score for every model

## 4.3 Discussion

### 4.3.1 Part 1: Data Exploratory Analysis

There are a total of 985 respondents were identified to respond to the questionnaire. 67.1% of the respondents are female, followed by male at 32.9%. The respondents are all grouped into certificate/diploma students (13.6%), undergraduate students (84.7%), postgraduate students (0.5%), master students (0.9%) and PhD (0.3%). Most of the respondents were 156 from the Computer Science/Information Technology course with 15.8% and followed by Architecture and Building with 107 respondents, Engineering with 92 respondents and 630 respondents with the percentage of 10.9%, 9.3% and 64.0%, respectively. This survey has been distributed across many institutions, and the highest number of respondents were from the University of Malaya which represents half of the respondents, 56.3%. The majority of respondents were Malaysians, which constitutes 98.4%, and the remaining were foreign students from Bangladesh, India, Somalia, Egypt, Sudan, Indonesia and China. In terms of household income, 59.6% fall under less than the RM4,849 income bracket and only 9.0%% of respondents with the highest income bracket, more than RM10,960. Based on the results in the learning mode preference, it shows that most of the students preferred face to face in their learning mode which consists of 79.2%, even for multiple answers, the mode is still within the choice among the respondents. The majority of the respondents had multiple social media accounts, and Youtube is the most popular social media platform among them, followed by Instagram, which contributes 70.9% and 56.2%, respectively. Whatsapp has the highest percentage compared to the other channel, about 89.1%. Whatsapp also exists in the other top percentage for students with the multi-answer. In addition, most of the students agreed that the main constraints in their online learning are the technical issue, followed by adaptability and time management. The findings of the preferred learning object revealed that most of the students opt for lecture notes as their learning object. Apart from those findings, the majority of the students answer Live Lecture for the technical or hands-on subject in terms of online instructional or assessment preference and followed by a demonstration. Most of the students agreed and were aware of their learning style in their learning process. They are familiar with the terms, aware and understand the degree of importance of learning style in their learning process.

### 4.3.2 Part 2: Data Modelling

From the previous section, this research project has witnessed how the three models fare in predicting the instructional strategy based on the accuracy rate across the three models. Although the second model, which is trained using NB, it has the lowest accuracy compared to SVM and GB, with the rate stood at 68.95%, but the training time took run this model is the shortest among the three models used in this research project. As mentioned in the literature, NB works well if the training set is small. Since this research has quite enormous number of rows and columns, as a result, this probably be the contributed reason NB cannot perform as good as the others.

Conversely, GB requires a lot more time to train when the data set is vast and high dimensional and thus causing it the longest among the three models used in this research project, albeit GB having better accuracy rate of 70.70% as compared to NB. This rings true to the literature review as GB is indeed the most cost-intensive algorithm among the three. Since our dataset is relatively small in size, GB probably is not a good option when the dataset grows in size and complexity in future.

Last of all, SVM has the highest rate of accuracy, which is 72.04% and the time spent for training is somewhat close to NB. The higher the validation metrics, the better is the classifier. As observed, all the validation metrics, including accuracy, precision and F1 score all have high values for SVM; thus, we can conclude that the classifier used in the approach has well performed. For this stated reason, most instructional strategy classifiers have move on to use Machine Learning techniques because of its high accuracy and the ability to classify the learning object based on personalisation. This means that when an educator wants to find a homogeneous instructional strategy as an illustration, the SVM model is able to understand the context and provide similar meanings for the corresponding learning style.

Table 4.2: Model Evaluation Table

| **Model** | **Accuracy (%)** | **Training Time** | **Remarks** |
|---|---|---|---|
| Gradient Boosting | 70.687970 | Longest | Costly |
| Naïve Bayes | 68.946701 | Shortest | Lowest Metric |
| Support Vector Machine | 72.043147 | Fairly Short | Best Result |

In conclusion, SVM is the best model in this study due to its superior accuracy, precision and F1 score. The result from this research is beyond the expectation based on the literatures considering how complex the dataset is by having high dimension and how massive it is by having so many features.

## 4.4 Summary

In this chapter, it gets started with exploratory data analysis from the result of the survey questionnaire. It also presented the findings into visualisation and with some insight in the subsequent discussion. Then the split of data into training and testing data to run the model based on three machine learning techniques, namely GB, NB and SVM. The model is then evaluated in the discussion to identify the best model for the classification of learning object based on personalisation. As a conclusion, SVM is found to be the best model in this study.

# CHAPTER 5: CONCLUSION

## 5.1 Introduction

The intention of this research project contemplates on utilisation of machine learning techniques and their applications in predicting instructional strategy. This chapter will then make a conclusion on the research project by revisiting the research objectives and answering the research questions mentioned earlier on. After that, a discussion on limitations also will be covered as well as the future work to be carried out.

## 5.2 Summary of Findings

Research Objective 1: To develop a classification model of online instructional strategies based on personalisation.

Research Question 1: How should the classification model of online instruction strategies based on personalisation be developed?

This research project achieved the first research objective for this research project, to be specific is to develop a classification model of online learning strategies based on personalisation. The classification model is developed by collecting the data from survey questionnaire on students' feedback and performed pre-processing method to clean the data as laid out in Chapter 3, from section 3.3 until 3.5. Personalisation was achieved by having learners' preference detected by VARK model as mentioned in Chapter 2, section 2.2.4. By doing data cleaning, we are able to get a better quality of data and generate features from the data. With these features, we can successfully classify the learning objects by personalisation by applying machine learning models listed in Chapter 3, section 3.5 namely GB, NB and SVM.

Research Objective 2: To evaluate the model performance of online instructional strategies based on personalisation.

Research Question 2: What is the performance of the classification model that is developed?

Supplementarily, this research project had also successfully achieved the second objective which is to evaluate the model performance of the instructional strategy classification by finding the best machine learning model in this research, as mentioned in Chapter 4, section 4.3.2. This research found SVM to be the best classifier model. This can be explained by SVM having the best score across the metric board among the other model.

## 5.3 Limitation

There are numerous limitations to the approaches used in this study. To begin with, the research is not comprehensive enough as it only selected three machine learning models for comparison while there are still other machine learning models which can be used and to be compared with their performance. For the purpose of improving learning capabilities, the study on instructional strategy and learning style should even more extensively be done across various machine learning algorithms. When it comes to data availability, the dataset size used is relatively small when classification of so many instructional strategies were done using only thirty personalisation feature data. Furthermore, this research project had analysed the instructional strategy to narrow down the target audience to university students only, which merely includes undergraduate and postgraduate students without primary or secondary students' penetration which should be covered holistically in the education system. The reason of why primary and secondary students was being excluded is to control the size and focus on the target criteria.

## 5.4 Future Work

To study more about instructional strategy among students in education system, students' feedback is highly recommended to acquire and strengthen the information. Moreover, this research project can maximise the learner's or student’s definition among primary and secondary students, which eventually can bring benefits to the entire education system as a whole in the future study. Besides, this could possibly give us some new insights as the group of people who are responding to the questionnaire are different. Apart from that, suggestion of other classification models is welcomed as well for similar research. Since this research has proven the SVM is the best model to classify the instructional strategy, future researchers can use this model as a benchmark to compare with the other popular models, such as Neural Network, which had not been studied in this study.

## 5.5 Conclusion

This research project intends to discover the potential of using machine learning to complement the conventional method in classifying instructional strategies in the education system. This research project compares and evaluates the performance of all three models and conclude that the result is more inclined to SVM to be the most promising approach in predicting instructional strategy based on personalisation. To the best extent, this is the first study incorporating student feedback in the research for instructional strategy, which assists in closing the gap with previous work done by others. The approach of this research project will have a substantial influence and mutual benefit between learners and educators.

## REFERENCES


Akdeniz, C. (2016). Instructional strategies. In *Instructional process and concepts in theory and practice* (pp. 57-105). Springer, Singapore.Danks, S. (2011). The ADDIE model: Designing, evaluating instructional coach effectiveness. *ASQ Primary and Secondary Education Brief*, *4*(5), 1-6.

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. John Wiley & Sons.

Ayodele, T. O. (2010). Types of machine learning algorithms. *New advances in machine learning*, *3*, 19-48.

Cai, W., Wei, R., Xu, L., & Ding, X. (2021). A method for modelling greenhouse temperature using gradient boost decision tree. *Information Processing In Agriculture*.

Catal, C., & Nangir, M. (2017). A sentiment classification model based on multiple classifiers. *Applied Soft Computing*, *50*, 135-141.

Cennamo, K., & Kalk, D. (2019). *Real world instructional design: An iterative approach to designing learning experiences*. Routledge.

Cheng, X., Ma, X. Y., Luo, C., Chen, J., Wei, W., & Yang, X. (2021). Examining the relationships between medical students' preferred online instructional strategies, course difficulty level, learning performance, and effectiveness. *Advances in Physiology Education*, *45*(4), 661-669.

Creswell, J. W. (2014). Qualitative, quantitative and mixed methods approaches.

Dhawan, S. (2020). Online learning: A panacea in the time of COVID-19 crisis. *Journal of Educational Technology Systems*, *49*(1), 5-22.

Drljača, D., Latinović, B., Stanković, Z., & Cvetković, D. (2017). Addie model for development of e-courses. In *Documento procedente de la International Scientific Conference on Information Technology and Data Related Research SINTEZA [Internet]* (pp. 242-247).

Eristi, B., & Akdeniz, C. (2012). Development of a scale to diagnose instructional strategies. *Contemporary Educational Technology*, *3*(2), 141-161.

Estrela, D., Batista, S., Martinho, D., & Marreiros, G. (2017, April). A Recommendation System for Online Courses. *In World Conference on Information Systems and Technologies (pp. 195-204).* Springer, Cham.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American journal of theoretical and applied statistics*, *5*(1), 1-4.

Fleming, N., & Baume, D. (2006). Learning Styles Again: VARKing up the right tree!. *Educational developments*, *7*(4), 4.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational statistics & data analysis*, *38*(4), 367-378.

Gamalel-Din, S. A. (2012, October). An intelligent etutor-student adaptive interaction framework. In *Proceedings of the 13th International Conference on Interacción Persona-Ordenador* (pp. 1-8).

Ghallabi, S., Essalmi, F., & Jemni, M. (2013, October). Toward the reuse of E-Learning personalisation systems. In *Fourth International Conference on Information and Communication Technology and Accessibility (ICTA*) (pp. 1-3). IEEE.

García-Gonzalo, E., Fernández-Muñiz, Z., García Nieto, P. J., Bernardo Sánchez, A., & Menéndez Fernández, M. (2016). Hard-rock stability analysis for span design in entry-type excavations with learning classifiers. *Materials*, *9*(7), 531.

Gregory, G. H., & Chapman, C. (2012). *Differentiated Instructional Strategies: One Size Doesn' t Fit All*. Corwin press.

Heinich, R., Molenda, M., Russell J. D. & Smaldino, S. E. (2002). *Instructional media and technologies for learning*. New Jersey: Merrill Prentice Hall.

Johnston, C. A. (1994, September). Unlocking the will to learn: Identifying a student's unique learning combination. In *British Educational Research Association Conference. Oxford Sept. 8th*.

Kaya, G., & Altun, A. (2011, October). A learner model for learning object based personalised learning environments. In *Research Conference on Metadata and Semantic Research* (pp. 349-355). Springer, Berlin, Heidelberg.

Kim, R. H., Gilbert, T., & Ristig, K. (2015). The effect of surgical resident learning style preferences on American Board of Surgery In-training Examination scores. *Journal of Surgical Education*, *72*(4), 726-731.

Li, T., Hu, S., Beirami, A., & Smith, V. (2021, July). Ditto: Fair and robust federated learning through personalisation. In *International Conference on Machine Learning* (pp. 6357-6368). PMLR.

Madeh Piryonesi, S., & El-Diraby, T. E. (2021). Using machine learning to examine impact of type of performance indicator on flexible pavement deterioration modeling. *Journal of Infrastructure Systems*, *27*(2), 04021005.

Mahmood, S. (2021). Instructional strategies for online teaching in COVID-19 pandemic. *Human Behavior and Emerging Technologies*, *3*(1), 199-203.

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & de Mendonça, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC research notes*, *4*(1), 1-14.

McTighe, J., & O'Connor, K. (2005). Seven practices for effective learning. *Assessment*, *63*(3).

Molenda, M. (2003). In search of the elusive ADDIE model. *Performance improvement*, *42*(5), 34-37.

Nilson, L. B., & Goodson, L. A. (2018). *Online teaching at its best: Merging instructional design with teaching and learning research*. John Wiley & Sons.

Oakleaf, M., & VanScoy, A. (2010). Instructional strategies for digital reference: methods to facilitate student learning. *Reference & User Services Quarterly*, 380-390.

Okano, K., Kaczmarzyk, J. R., & Gabrieli, J. D. (2018). Enhancing workplace digital learning by use of the science of learning. *Plos one*, *13*(10), e0206250.

Popescu, E. (2010). Adaptation provisioning with respect to learning styles in a Web-based educational system: an experimental study. *Journal of computer assisted learning*, *26*(4), 243-257.

Sadler, D. R. (2014). Beyond feedback: Developing student capability in complex appraisal. In *Approaches to assessment that enhance learning in higher education* (pp. 55-70). Routledge.

Srivastava, N., & Schrater, P. (2015). Learning what to want: context-sensitive preference learning. *PloS one*, *10*(10), e0141129.

Suartama, I. K., Setyosari, P., & Ulfa, S. (2019). Development of an instructional design model for mobile blended learning in higher education. *International Journal of Emerging Technologies in Learning*, *14*(16).

Sun, Z., Bebis, G., & Miller, R. (2004). Object detection using feature subset selection. *Pattern recognition*, *37*(11), 2165-2176.

van Geel, M., Keuning, T., Frèrejean, J., Dolmans, D., van Merriënboer, J., & Visscher, A. J. (2019). Capturing the complexity of differentiated instruction. *School effectiveness and school improvement*, *30*(1), 51-67.

Wong, T. T., & Tsai, H. C. (2021). Multinomial naïve Bayesian classifier with generalised Dirichlet priors for high-dimensional imbalanced data. *Knowledge-Based Systems*, *228*, 107288.

Yang, Z. (2022). Data analysis and personalised recommendation of western music history information using deep learning under Internet of Things. *PloS one, 17*(1), e0262697.

Yavgildina, Z., & Mishina, A. (2015). Principles of personalisation of content of the competency-based training of students of the artistic-pedagogical specialisation.