CLASSIFICATION OF LEARNING OBJECT BASED ON PERSONALIZATION

SITI AZIRA BINTI SUHOT

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

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

2021

UNIVERSITY OF MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: SITI AZIRA BINTI SUHOT (I.C/Passport No: 91112701-5750)

Matric No: WQD170083/ 17058066/1

Name of Degree: MASTER OF DATA SCIENCE

Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"): CLASSIFICATION OF LEARNING OBJECT BASED ON PERSONALIZATION

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: 18 June 2021

Subscribed and solemnly declared before,

Witness's Signature Date: 18 June 2021

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

Name: Nor Liyana Binti Mohd Shuib

Designation: Assoc Prof Dr

CLASSIFICATION OF LEARNING OBJECT BASED ON PERSONALIZATION

ABSTRACT

Personalization is important in the learning process. If educators know and understand student's persona, they can propose a suitable presentation of materials or learning object, which eventually can drive the students to understand the subject better and enhance their learning capabilities. Educators tend to provide various types of learning objects to cater to different type of personalization. A learning object is a tool for learning to facilitate a deeper understanding of the information. In fact, researchers have proposed various educational tools such as e-learning. In the process of recommending learning objects to students, researchers need to identify which learning objects belong to which personalization. Moreover, most of the researchers do not address the classification of learning objects based on students' personalization. Most of them classify the learning objects based on the description of learning style preference itself. Hence, this research proposed a classification of learning objects using supervised machine learning by gathering students' feedback to cater for the gap in the previous work. A survey is conducted to gain the dataset. The dataset is preprocessed and then modelled using Naïve Bayes, Decision Tree and Support Vector Machine. The result of the study is able to provide an important insight into the education sector, which eventually will mutually beneficial to both educators and learners.

Keywords: Learning object; personalization; machine learning

KLASIFIKASI OBJEK PEMBELAJARAN BERDASARKAN KEPERIBADIAN

ABSTRAK

Keperibadian penting dalam proses pembelajaran. Sekiranya pendidik mengetahui dan memahami keperibadian pelajar, mereka boleh mencadangkan persembahan bahan atau objek pembelajaran yang sesuai, yang akhirnya dapat mendorong pelajar memahami subjek dengan lebih baik dan meningkatkan kemampuan belajar mereka. Pendidik cenderung menyediakan pelbagai jenis objek pembelajaran untuk memenuhi perbezaan dalam jenis keperibadian. Objek pembelajaran adalah alat untuk belajar untuk memudahkan pemahaman maklumat yang lebih mendalam. Sebenarnya, penyelidikan telah mencadangkan pelbagai alat pendidikan seperti e-learning. Dalam proses mengesyorkan objek pembelajaran kepada pelajar, penyelidik perlu mengenal pasti objek pembelajaran yang menjadi kepunyaan. Lebih-lebih lagi, kebanyakan penyelidik tidak menangani klasifikasi objek pembelajaran berdasarkan keperibadian pelajar. Sebilangan besar daripada mereka mengklasifikasikan objek pembelajaran berdasarkan perihalan keutamaan gaya pembelajaran itu sendiri. Oleh itu, penyelidikan ini mencadangkan klasifikasi objek pembelajaran menggunakan pembelajaran mesin yang diselia dengan mengumpulkan maklum balas pelajar untuk mengatasi jurang dalam karya sebelumnya. Sebuah kaji selidik telah dijalankan untuk mendapatkan kumpulan data. Kumpulan data tersebut telah menjalani proses terdahulu and dimodelkan dengan menggunakan model seperti *Naïve Bayes*, *Decision Tree* dan *Support Vector Machine*. Hasil kajian ini dapat memberikan pandangan penting dalam sektor pendidikan yang akhirnya akan saling memberi manfaat kepada pendidik dan pelajar.

Kata Kunci: Objek pembelajaran; keperibadian; pembelajaran mesin

ACKNOWLEDGEMENTS

Syukur, Alhamdulillah. All the praise goes to Allah for giving me the ability, patience and strength to finally complete this research project. I would like to express my profound appreciation and gratitude to my research project supervisor, Assoc Prof. Dr. Nor Liyana Binti Mohd Shuib of the Faculty of Computer Science and Information Technology at University of Malaya. Each and every question and guidance that I had and received regarding my research project was duly accommodated by Dr. Yana without hesitation. I would also like to thank to all the respondents of this study for their kind cooperation and contribution. Finally, my deepest appreciation to my parent and parent in law for the support, prayer and encouragement throughout the year of my study and research. And more than anyone else with his unconditional love, my husband, Mohd Shahzerin, for keeping the support and comfort and being understanding while things run simultaneously at this point of time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

LO	:	Learning Object
LS	:	Learning Style
SVM	:	Support Vector Machine
DT	:	Decision Tree

CHAPTER 1: INTRODUCTION

1.1 Research Background

Research shows that every student preferred to use training material based on their own unique learning style, interests and needs. It is widely accepted that understanding students' learning style and preferences can be very beneficial for both educators and students. Presentation of materials that are parallel with students' learning style preference can enhance the learning process. (Normadhi et al., 2019). Educators tend to provide various types of learning objects to cater to different type of learning style. A learning object is a tool for learning (Smith, 2004) to facilitate a deeper understanding of the information (Franzoni & Assar, 2009). There is a tendency to claim that a personalized learning experience such as intelligent tutoring systems, adaptive educational hypermedia systems, adaptive educational systems and semantic web-based education systems can cater to various learning objects. Personalized learning removes location, time and other limitations in the teaching process and aims to tailor teaching for each learner's constantly changing needs and skills (Sampson, Karagiannidis & Kinshuk, 2002). In other term, personalization is defined as adapting learning experience to different students based on the analysis of knowledge, skills, interests and preferences of individuals (Davedzic, 2006). In traditional learning, materials are prepared for an average learner. But in personalized learning environments, materials can be adapted based on academic records, psychological variables, skills and learning environment preferences of the learner. The content can be made more complex or simpler according to the needs or demands of the learner (Martinez, 2010).

The use of learning objects for educational purposes brings various mileage in content knowledge and preparing educationally sound content and delivery. Having considered the advantages that learning objects can bring, the classification of learning objects based on personalization could not be found in the reviewed literature. For this reason, in this research project, a classification of learning object is proposed for learning system that is based on students' personalization.

1.2 Problem Statement

The learning style of a learner is identified by the way in which the information is being received and processed. In traditional ways, predefined mathematical equations and a set of questionnaires are used to determine the learning style. This approach may not be appropriate due to student's preference may have more than one mode of learning style. To gratify a given learning style, the educator must use the approach that could meet the demands and needs of diverse learning perspective.

Finding an appropriate learning material is tough for student and almost always, they cannot find the learning style that suits their needs and requirements (Shuib, 2013; Normadhi et al., 2019). Most of the researchers classify the learning objects based on the description of learning style preference itself. A paper by Fleming described learning styles as individuals' characteristics and preferred ways of gathering, organizing and thinking about information (Fleming, 2005). Moreover, there is insufficient literature on considering students' feedback in previous research. The feedback from students should be considered during the evaluation process to address more appropriate learning objects to students (Shuib, 2013; Balasubramanian & Anouncia, 2018). Besides learning styles, personalization can be delivered based on information concerning students' level of knowledge (Laksitowening et al., 2017) . In terms of learning object, it can be classified according to its difficulty level, then subsequently it will be presented based on students' ability level (Kim, Jung, Lim & Kim, 2009). Hence, there is a need to explore the classification of learning objects by exploring the concept of personalization of students to improve the quality and effectiveness of the learning and education system.

1.3 Research Objectives and Research Questions

This research project desires to gauge the potential of using the classification technique of learning objects based on student's personalization. The study aims to address the following research objectives:

Objective 1: To develop a classification model of learning object based on personalization.

Objective 2: To evaluate the model performance of learning object based on personalization.

These are the research questions corresponding to the objectives that we intend to answer:

Question 1: How should the classification model of learning object based on personalization be developed?

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

1.4 Research Method

The general methodology of this study consists of five phases and is shown in Figure 1.1 as follows. In the subsequent section, each phase will be explained to briefly discuss the core contents. Detailed discussion is performed accordingly in Chapter 3.

Figure 1.1: Research Methodology

1.4.1 Problem Identification

This phase identifies the problem through literature reviews on the classification of learning object based on personalization. The outcome of the reviews is discussed thoroughly in Chapter 2.

1.4.2 Data Collection

This step deliberates on the study setting and dataset used. A questionnaire has been distributed accordingly within the research period to collect the input from students. The details for this step are presented in-depth in Chapter 3.

1.4.3 Data Pre-Processing

This step is being carried out in order to ensure a robust and high-quality dataset. This includes data cleaning, feature selection, handling imbalanced data and data standardization. The details are further concentrated and elaborated in Chapter 3.

1.4.4 Model Development

This step uses a classification algorithm/technique to classify the learning object based on personalization. The discussion of model development is described entirely in Chapter 3.

1.4.5 Model Evaluation

This step will evaluate the performance of the model through the confusion matrix and area under the curve (AUC). Ultimately, this step identifies the best classification model for learning object among students. A more comprehensive discussion is done in Chapter 3 and 4.

1.5 Research Scope

The present work attempts to utilize the machine learning approach to classify learning object based on personalization. The scope for this research project will be as follows:

- Only undergraduate and postgraduate students are selected for the purpose to achieve the objective in this study. Hence, the research material is collected and confined to university students.
- An extensive data cleaning to qualify the dataset as high quality is performed.
- Prediction of the learning object using classification algorithms. A total of three algorithms are employed throughout this research project, namely Naïve Bayes, Support Vector Machine (SVM) and Decision Tree.

1.6 Research Significant

The significance of this research was observed mainly in the education domain. Currently, no learning style that considers students' feedback in the evaluation process. The involvement of students' persona is a new concept in classifying learning object. The approach of machine learning is rather new in the local education setting. The data is collected purely from the targeted and actual respondents and make the data valuable and priceless. Through extensive data cleaning, the model developed shows confidence result and can serve as the platform or benchmark for a machine learning approach in the educational environment that may enhance learning capabilities between educators and learners in the near future.

1.7 Organization of Research

This research project is structured into five chapters:

Chapter 1	This chapter will include the research background, which will serve as the introduction and overview of this study. It is then followed by the statement of problem in which the domain problems are being addressed. The research objectives and questions are listed, and the research method is clearly itemized. The limitation and significance of study are being included as well. The structure of research is elaborated in this chapter too.
Chapter 2	This chapter will briefly discuss the literature review for learning and its content, personalization, classification and the different approached that have been conducted in previous studies. The summary is being done based on the learner model used in each paper.
Chapter 3	This chapter will examine the research methodologies used for each of the objective intended in this project. We will begin with the research design framework, providing the overview of how we can address this study. A detailed explanation of data acquisition and data cleaning will be discussed as well. Lastly, we will talk about the model that will be developed to materialize the research objectives.
Chapter 4	This chapter will discuss the results and findings of the research objectives. We will discuss about the model evaluation, data analysis and

interpretation. We will also explore the possible insights that can be gathered in the findings.

Chapter 5 This chapter will conclude the research project with an overview of what have been done to achieve its objectives. The contribution, limitations and future work recommendation of the study is presented accordingly in this chapter.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

The goal of this chapter is to initiate the major terms used throughout this research project. This chapter is literally divided into four main sections. The first section introduces the background of learning. The second section reviews student personalization. Next, the third sections explore the previous work done in learning object prediction using the machine learning method as well as the discussion on the limitations of the applications. The last section in this chapter will portray the framework of this study.

2.2 Learning

Learning can be defined as a switch of behaviour as out-turn of experience which subsequently will increase the potential to improve the performance and future learning. It is the process that involves reception and transformation once information is received. At the reception phase, diverse senses are captivated in collecting information from external sources while transformation, it is resulted from internal sources such as memorization, inference, pondering, inception as well as reflection. A learner is a person who is still learning something, and its learning process is continuous. In this research project, the learner refers to the student. Every student has uneven learning needs to acquire and process the gained knowledge. If the student can understand their learning needs and materials that suit and meet their needs, it will be very helpful for them to improve their performance and learning.

2.2.1 Learning Style

There are relative variables to identify the learning style of a learner. Learning styles can be described as students' preferred ways to learn (Truong, 2016). Besides that, it refers to the preferential way in which the student perceives, processes, understands and retains information (El Aissaoui et al., 2019).

Learning style is explicitly significant in the learning process. Learning style carried a lot of different terms in literature, such as cognitive style, personality types and some might use sensory preferences. In some scenario, some of the terms have been used mutually, while in other instances, they have been distinguished (Cassidy, 2004). As such, learning style can be explained as the complex manner in which and conditions under which students most adequately perceive, process, recall and store materials that they intend to learn (James & Gardner, 1995). While for another paper by Reid, the term cognitive style is being used with definition as an individuals' natural, pattern and preferred way of captivating, processing and reserving new skills and information (Reid, 1995). Mortimore (2003) has distinguished the term between learning style and cognitive style. He explained that learning styles could be viewed more in terms of strategies to deal with hand and considered it to be less stable. On the other hand, he emphasized that cognitive style is relatively stable. However, there is no strong justification to distinguish between learning style and cognitive style as some authors have been using cognitive style in a more generic way that includes learning style (Williamson & Watson, 2006). It relies upon the psychological, cognitive, environment as well as prior experience of an individual. According to Peterson, learning styles constitutes various inter-related components including instructional preferences, modalities, and learning strategies and the used of these elements can be influenced by the environment and type of learning task (Peterson, 2009). On the other way, learning style can be referred to as the way in which a learner or student is collects, processes, understands and retains the information in a single or combination of ways which eventually affects the outcomes of learning (Zapalska & Brozik, 2006).

2.2.1.1 Learning Style Model

Research on learning style has evolved since the early 1960's (Howles, 1960). There are huge volumes of learning style model in the literature to classify the learning styles. The popular learning models will be discussed thoroughly in the following section.

1) David Kolbs' Learning Style Model

This learning style is based on the Experiential Learning Theory (Kolb, 1984;1999). The model comprised two levels; the first level consists of four-stage, and all stages must be integrated and engaged in making effective learning. While for the second level, the mix of two preferred stages is represented. The model is measured using learning style inventory (LSI).

In the first level of Kolb's learning style model, there are four stages:

- Concrete Experience (CE) – At this stage, the learning occurs through exposure or situations encountered or modification of current exposure.
- Reflective Observation (RO) – Here, learning happens after gaining experience that allows the learner to discuss and ask any questions.
- Abstract Conceptualization (AC) – This stage will enable the learner or student to acquire new knowledge or a modification of an existing theoretical concept.
- Active Experimentation (AE) – For this stage, the learner or student will test their knowledge in the real world and gain new experience.

Meanwhile, the second level of the model can be illustrated and explained as follows:

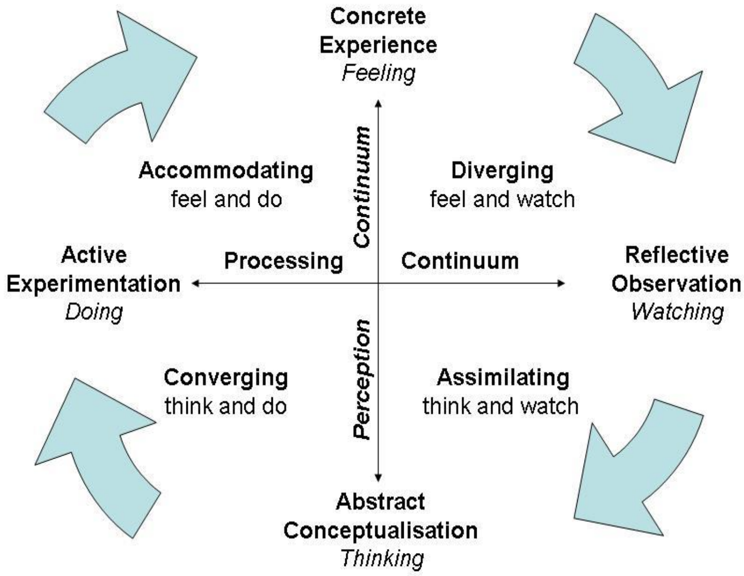


Figure 2.1: Kolb's LS (Kolb, 1999)

- Diverging (CE/RO) – Basically, diverging consists of two stages, namely concrete experience and reflective observation. At this stage, learner or students are imaginative and creative with ideas and also capable of seeing a situation from different angles or perspectives.
- Assimilating (AC/RO) – Assimilators are described as a combination of abstract conceptualization and reflective observation. Learner or student in this category is competent in creating theoretical models according to inductive reasoning.
- Converging (AC/AE) – For converges, it consists of abstract conceptualization and active experimentation stage. The advantage is skewed in making practical applications of ideas and to work out with problems, converge learner or student is using deductive reasoning.
- Accommodating (CE/AE) – Learners or students who are accommodators uses concrete experience and active experimentation. They are actively engage with the world and executing things instead of studying and reading only.

2) Felder-Silverman Learning Style Model

Following to Felder and Silverman, in order to make an impactful learning process there is a need to combine learning elements. The elements might include visual against verbal, sensing against intuitive, deductive versus inductive, reflective against active and sequential against global (Felder & Silverman, 1988; Felder, 1995). Unlike Kolb's Model, this model is measured using the Index of Learning Style (Graf et al., 2007). Previous research has discussed that this model is the best approach for engineering course (Honey & Mumford, 1992).

3) Honey and Mumford Learning Style Model

The learning style developed by Peter Honey and Alan Mumford is evolved from Kolb's learning style for use in commerce (Honey & Mumford, 1992). They have discovered four definite learning style, including reflectors, theorists, activists and pragmatists. It is based on the learning cycle. Hone and Mumford configured a set of questions that can be used to support the identification of learner's learning style that constant in nature.

4) Dunn and Dunn Learning Style Model

This learning style is suggested to be one of the oldest and most widely used. It assesses individuals concerning global approaches to learning (Dunn & Dunn, 1978). In accordance with Dunn and Dunn, the learning style can be categorized into five elements (Carma, 2013).

- Environmental – Hotness, light, sound and seating arrangements are examples of environment that can influence the learning style.

- Emotional – Emotional can be described as responsibility, motivation and perseverance and motivation related.
- Sociological – This element focusses on identifying the preference of the learning environment and learning in pair.
- Physiological – Physiological is mainly about how a learner responds to the learning tasks. It establishes other learning styles such as kinesthetic, auditory and visual.
- Psychological – Psychological emphasized about how a learner or student process and respond to the information and knowledge received.

5) VARK Model

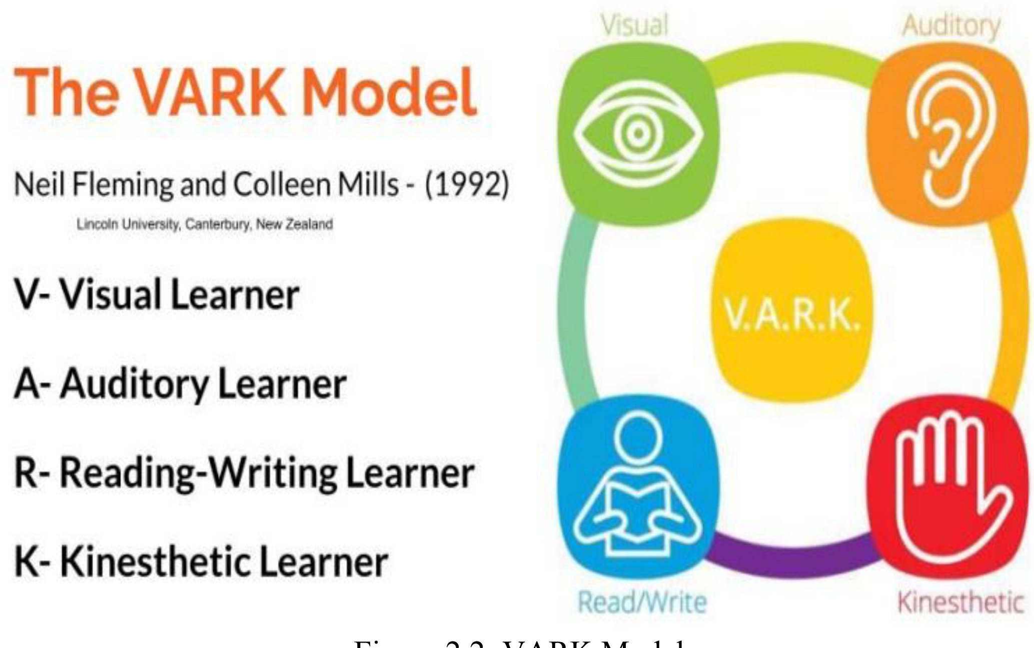


Figure 2.2: VARK Model

VARK stands for Visual, Auditory, read or Wrote and Kinesthetic, which representing learning object. The following table will provide a rundown about the learning style and learning approach based on the VARK model.

Table 2.1: Summary of VARK Model

No.	Learner Style	Learning approach towards learning	Learning Content
1.	Visual	Knowledge and understanding are acquired through the maps, images, diagrams and graphic representations	Power point slides, URL and video
2.	Auditory	Understanding content through discussion, speaking and listening	Power Points slides and URL with audio or recorded notes
3.	Read/Write	Learning through text or printed words to obtain information	Power Points slides, PDFs and text file
4.	Kinesthetic	Learning preference through practical, real time example, project as well as hands on activity. The learner or student learn by holding, moving, feeling and touching something	Case studies, demonstrations and hand on practice

In this research project, VARK model proposed by Neil D. Fleming is chosen. It is evolved from Gardner's's theory of multiple intelligence (1993). This model has been tested and validated empirically by Leite et al. (2009). The reason for choosing this model is mainly because:

- The model is one of the best models to achieve the objective of this paper, to classify the learning style.
- VARK questions are due to a real-life scenario. Hence respondents can relate to the questions easily (Rogers, 2009).
- Respondents have the privilege to answer according to their preference by answering more than one answer for every question, thus can provide information if they belong to multiple learning modes (Slater et al., 2007).
- Four Learning Style preferences represent in VARK corresponded accordingly with eight intelligence dispositions identified by Gardner and outlined by Silver (Gardner, 1999; Silver et al., 2000).

- To understand and explain learner's preference to learn, VARK is an accessible method to be used.

2.2.2 Learning Object

Due to the various learning style of a learner, hence it is important to develop and identify strategies to meet the student preferences in delivering learning object either physical or virtual. In terms of learning object, it can be explained as an any digital entity such as a text, a movie, an animation, an instructional content or a composition of all into a larger object according to defined educational purpose (Macedo & Ulbricht, 2012). They added that it could be located and reused, either stand-alone or in a composition of larger objects with defined goals and educational strategies in a different context.

Learning objects can be detailed as any form of a digital resource that is being used in order to carry out a learning activity. For instance, multimedia content, web page references, textual content, visuals, demonstration and many others. It is aimed to locate content from the web and reuse it in various educational environments (Macedo & Ulbricht, 2012). In such, the learner collects the content materials and used technology to assist the learning process. In a paper by Kolb, the characteristics of learning object include time span, size, reusability, interoperability and multiple contexts of the content (Kolb, 1984). To the same degree, time spent by a learner towards the particular learning content type is used to identify the importance of the learning object as technology evolves. A learning object can be classified based on difficulty level and subsequently be presented based on the learner's ability level.

2.2.3 Relationship between Learning Object and Learning Style

It is prevalent that in the circumstances where learning object is matched with individual learning styles, students tend to learn more effectively (Haq & Chand, 2012). The findings from a paper by Mestre reported that by connecting learning object and learning style, students are more effective in acquiring information (Mestre, 2010). In addition, he suggested to provide diverse learning object for all learning styles in order to assist students to learn in their preferred ways. Previous work done has studied the relationship between teaching strategies and learning styles and the result shown that the student responds differently to the teaching strategy according to their learning style, Hence, educators should be encouraged to teach according to students preference in learning style because of the tendency of teaching strategies that are more closely related to learning styles.

2.2.4 Personalized Learning

Along the process, personalization exists through the study of students' preferences. It can be expressed as an educational approach that aims to customize learning based on each students' strengths, needs, skills and interests. In personalized learning literature, personalization is divided into five groups following its complexity parameter (Martinez, 2000).

- Name-based personalization – Addressing user by name once the user logs into the system by username and password
- Self-described personalization – The system is taking user preferences, history and attributes by tools such as pre-tests, questionnaire or form.
- Segmented personalization – Grouping learners or students by demographic and common attributes such as class, course, department and etc.
- Cognitive personalization – Delivering content and teaching based on the cognitive process, preference, strategy and skill of learners or students. Next, the system will transform the content.

- Whole-person personalization – A combination of cognitive personalization and psychological resources that influence leaning and performance. The system will create an inference about the user in the learning process with constant update. Hence, user can be presented in all ways.

Most of the researcher's literature proposed whole-person personalization for personalization. The personalization enables the system to reflect the characteristics of learners (Ghallabi et al., 2013). The selection or customization of learning content and activities are being prepared for the need of learners' personalization. Personalization is commonly referring to the learning style, affective state and knowledge level of a learner. There are many learning styles models that can be applied for personalization as discussed in the previous section. Personalized can be conferred in various forms such as personalized learning environment (Gamalel-din, 2012), personalized learning content and personalized interfaces.

2.3 Review of Classification Algorithms

A classifier can be defined as an algorithm that takes a set of features that will categorize objects and uses them to determine the class of each object. Generally, in machine learning, there are two types called supervised and non-supervised machine learning (Zhang et al., 2013). Supervised machine learning, the machine was using a human expert to classify the object and provided a sample of objects with known classes. This set of knows objects is called the training set because it learns how to classify objects bases on classification programs. Unlike unsupervised machine learning, the machines work directly from the data and there is neither training nor pre-determined classes identified. The figure follows the steps involve in developing classifiers:

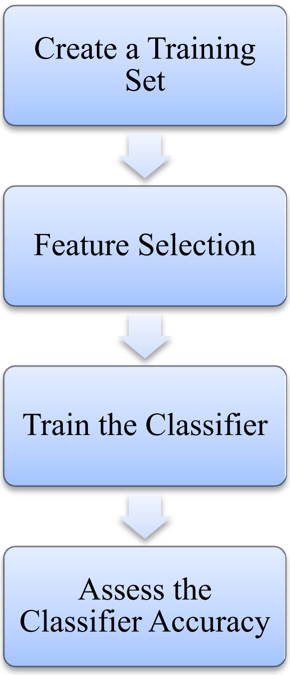


Figure 2.3: Object Classification

In machine learning, classification is a supervised learning approach where the computer learns from the training data that contains the variables and the labelled data. In a simpler way, classification refers to predictive modelling in which a class label is predicted for a given input data to draw some conclusion. Further section reviewed the past papers which applied classification algorithms that are used in this research project.

2.3.1 Naïve Bayes

Naïve Bayes is a supervised machine learning, and it is an extension of the Bayes theorem. There are two parts to this algorithm, namely Naïve and Bayes. The theorem gives an assumption of independence amid predictors. In other words, the classifier assumes that the existence of a particular feature in a class is independent to the existence of any other feature. Albeit the features are depending on each other, the properties that contribute to the probability is independent. The Naïve Bayes model is easy to develop and most common applied in comparatively large dataset. It has two simplifications in which the first simplification is to use conditional independence assumption and the second simplification is to ignore the denominator (Ouafae et al., 2018). Figure 2.4 shows the formula for the Bayes theorem in accordance to the Naïve Bayes theorem.

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

Figure 2.4: Naïve Bayes Theorem

2.3.2 Support Vector Machine

Support Vector Machine is an approach for the classification of both linear and non-linear data. Support Vector Machine is a type of supervised learning models and it associated with a learning algorithm which is used for data analysis and pattern recognition (Sun et. al., 2004). Basically, the Support Vector Machine takes input data, and for every given input, it will form an output of a non-probabilistic binary linear classifier. A paper by Ouafae et al. has used combining supervised and unsupervised machine learning algorithm to predict the learners' learning style, and the results show that the approach used is performing well (Ouafae et al., 2019). The SVM is called as a discriminative classifier that is formally designed by a separative hyperplane. It can be portrayed as points in space that are mapped; hence the points of distinct categories are separated by a gap as wide as possible. The aim of SVM is eventually to segregate the given data in the best possible way. Once it is done, the distance between the nearest points is being called a margin. The approach is to choose a hyperplane with the maximum margin between the support vectors in the dataset. The figure illustrated below is to visualize the Support Vector Machine.

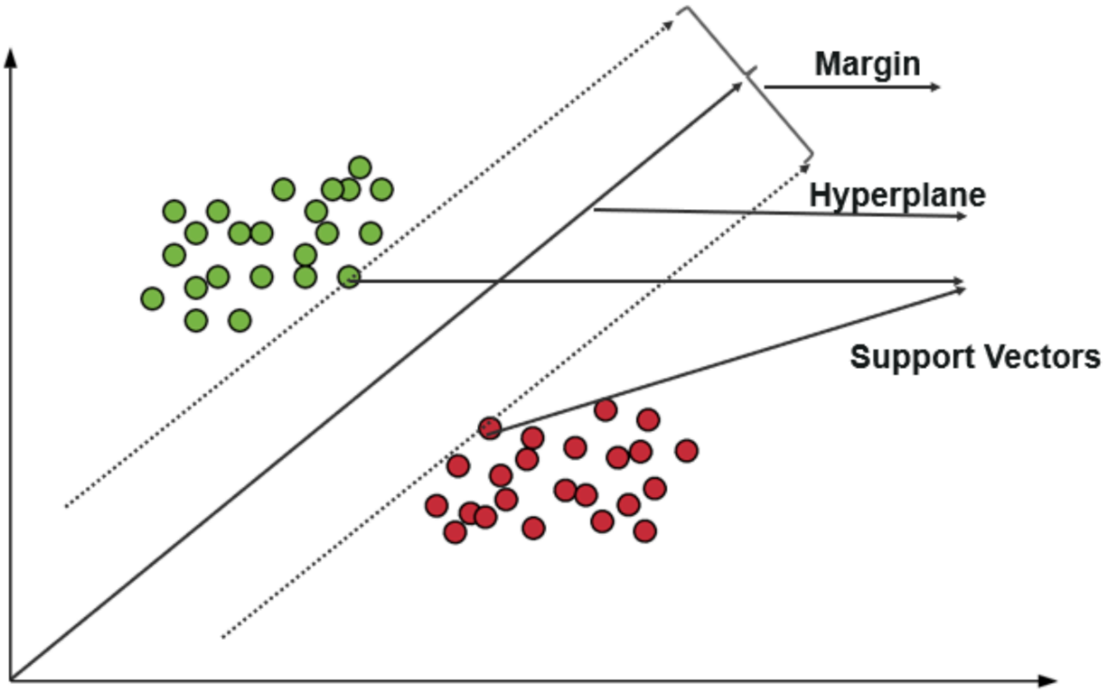


Figure 2.5: Support Vector Machine

2.3.3 Decision Tree

Decision Tree classifiers are regarded to be the most well-known methods in data classification (Bahzad & Adnan, 2021). It is supported by a paper that claimed Decision Tree as one of the powerful methods commonly used in various fields like machine learning, pattern identification and image processing (Stein et al., 2005). The model is uniting a series of basic test efficiently where a numeric feature is compared to a classification fulfilment, model development and followed by evaluation (Hilel et al., 2020). In the conceptual framework, the model is much easier to construct between the nodes. Hence, for grouping purpose, a Decision Tree is mainly used furthermore in model classification. Each tree composed the nodes and branches. Each node represents a feature in a group to be classified and each subset defined a value from the node (Dey, 2016). The figure below illustrated the structure of the Decision Tree.

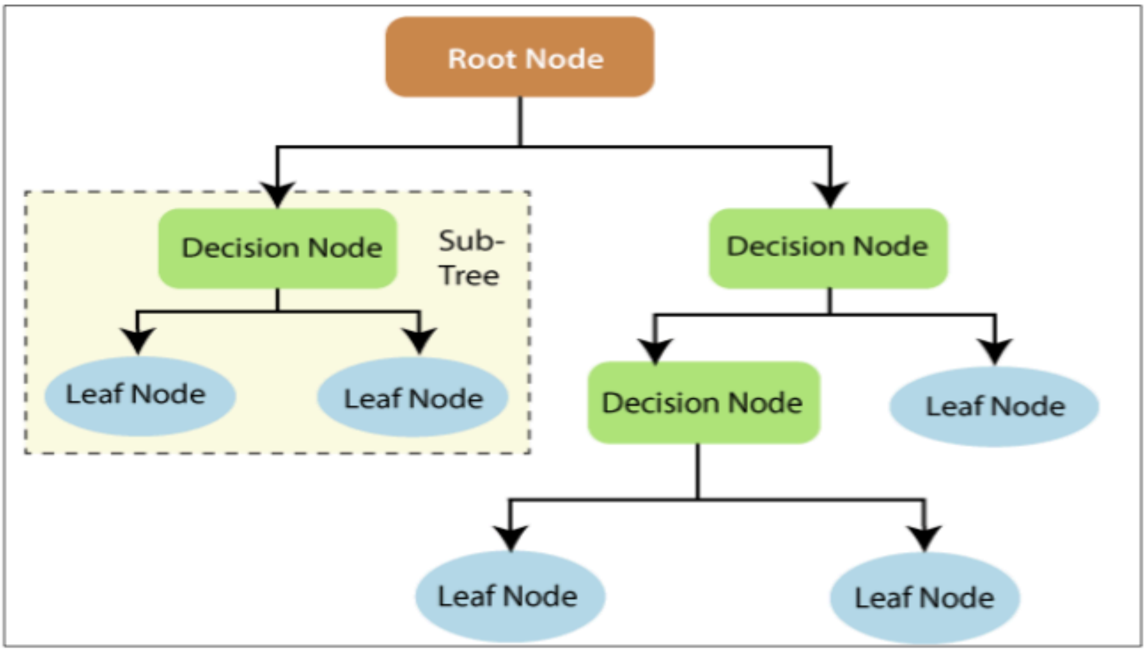


Figure 2.6: Structure of Decision Tree

2.4 Theoretical Framework

The Theory of Multiple Intelligence (Gardner, 1993) is based on the confidence that each individual has a distinctive intelligence or preferences. The VARK learning style model was evolved from this theory and chosen in this research project. The theoretical framework of this research project is constructed due to this theory and model, as shown in the following figure.

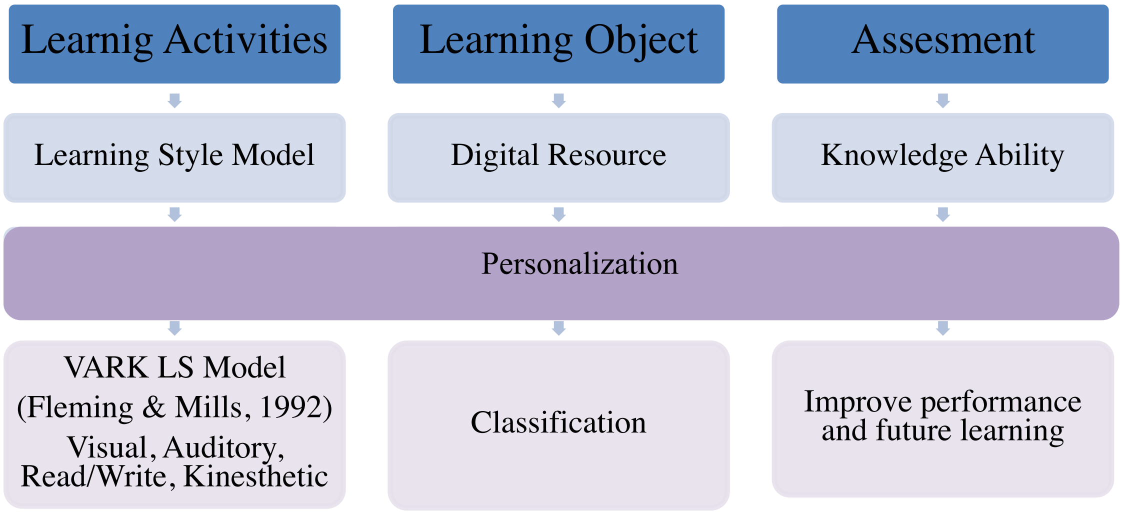


Figure 2.7: Theoretical Framework

2.5 Previous Works on Classification of Learning Object

The use of machine learning on learning object classification has been carried out in several studies. There are many researchers proposed in order to facilitate educators in matching learning object and student's preference in learning style. A study on learners classification for personalized learning experience in e-learning systems has used a prominent learning style called VARK to classify the learners (John et al., 2021). The study did not classify the learning object; instead classified the learner to recommend the learning object based on their learning preference. In addition, there is no machine learning involve in performing the classification. The study mainly used the vector of the percentage of learning objects and time spent by the learner on learning content. A study done by Bucker & Elaine (2017) discussed using gamification as a tool in teaching, learning and assessment. The study does not support the use of learning objectives to identify student preference in learning style. The literature believed in the development of gamification to be the continual renewal in the education system based on the positive impact of gamified learning on learning outcomes.

Shuib (2013) had written a paper by taking the text as a learning object. The study focused on developing information-seeking tool to identify student's learning style during their retrieval of information process. The study is limited to text merely and not taking image into consideration. The dataset is relatively considering low, and students' feedback is not being part of the study input. A recent paper by Biabani & Izadpanah (2019) have used KOLB learning style in the study. However, the study does not reflect a general result as the sample is among the Iranian only.

Meanwhile, a paper written by Stirling et al. (2017) has assessed nursing students had some preference on kinaesthetic learning compared to other attributes in VARK model. The study is using students' feedback in gathering the information, however, there is no predictive model to be applied from the information. The results are mainly discussed on the student's preference according to VARK learning style and by comparing the results with other nursing students.

Essalmi et.al. (2015) has concluded in considering different personalization strategies in the paper. The study is lack of having feedback about additional constraint to consider. Another study that perform about classification is done by Ouafae et. al. (2018) that conclude the approach yields to excellent result. The study is combining supervised and unsupervised machine learning approach to predict students learning style. The research literature is summarized in Table 2.2 as follows.

Table 2.2: Summary of Literature of Approach, Learning Model and Learning Object

Reference	Approach	Learning Style Model	Learning Object	Limitations
Martin et al. (2021)	Adaptive e-learning	VARK	-	• The classification is towards the learner • Not considering machine learning approach

Buckley & Elaine (2017)	Adaptive Learning	FSLSM	Gamified Learning Intervention	• Gamification should to be seen as a specific tool in the teaching, learning and assessment toolkit, and used as part of a holistic instructional design process
Shuib (2013)	Adaptive Learning	VARK	Text	• Feature Extraction Algorithm cannot extract vector images, symbols and size of image from PDF document • Sample size for the experiment • Not considering students' feedback
Biabani & Izadpanah (2019)	Adaptive Education	Kolb LSM		• A comprehensive study needed to be done on students from different years of studying English as a foreign language in order to provide more generalisable results
Stirling et al. (2017)	Adaptive Suggestion	VARK	-	• Not considering machine learning approach
Essalmi et al. (2015)	Adaptive E-Learning	FSLSM	Hypermedia	• Lack of having feedback about additional constraints to consider • Using different personalization strategies
Ouafae et al. (2018)	Adaptive e-Learning	FSLSM	Student behaviours	• Using log file of student's info instead of student's feedback

2.6 Summary

In this section, the topics were explicitly reviewed. Based on the literature review, problem statements and research problem was derived. The findings of this research suggest the importance of learning style in the learning process. Learning style plays a big role, especially for learner or student. While students acknowledge their own learning style, they will be able to merge it into their learning process and experience. Hence, the learning process and experience will be faster, easier, effective and successful. Furthermore, understanding learner style can assist the learner in how to learn, which resulted in the learner become accountable to their own learning. Indirectly, learners' confidence will be boosted and more encouraging. As an educator or teacher, this will help them to facilitate and design lesson planning that matches their student's preference. On the other occasion, mismatching should be handled with care as it will lead to learners' dropouts (Tuan, 2011). The next chapter discusses the methodology taken during this research.

CHAPTER 3: METHODOLOGY

3.1 Introduction

This chapter briefly described how the research project be tackled to achieve the objective and solve the problems in the problem statement. The methodology of classifying the learning object is illustrated in Figure 3.1 as follows, where the entire research design of this research project is divided into five main processes:

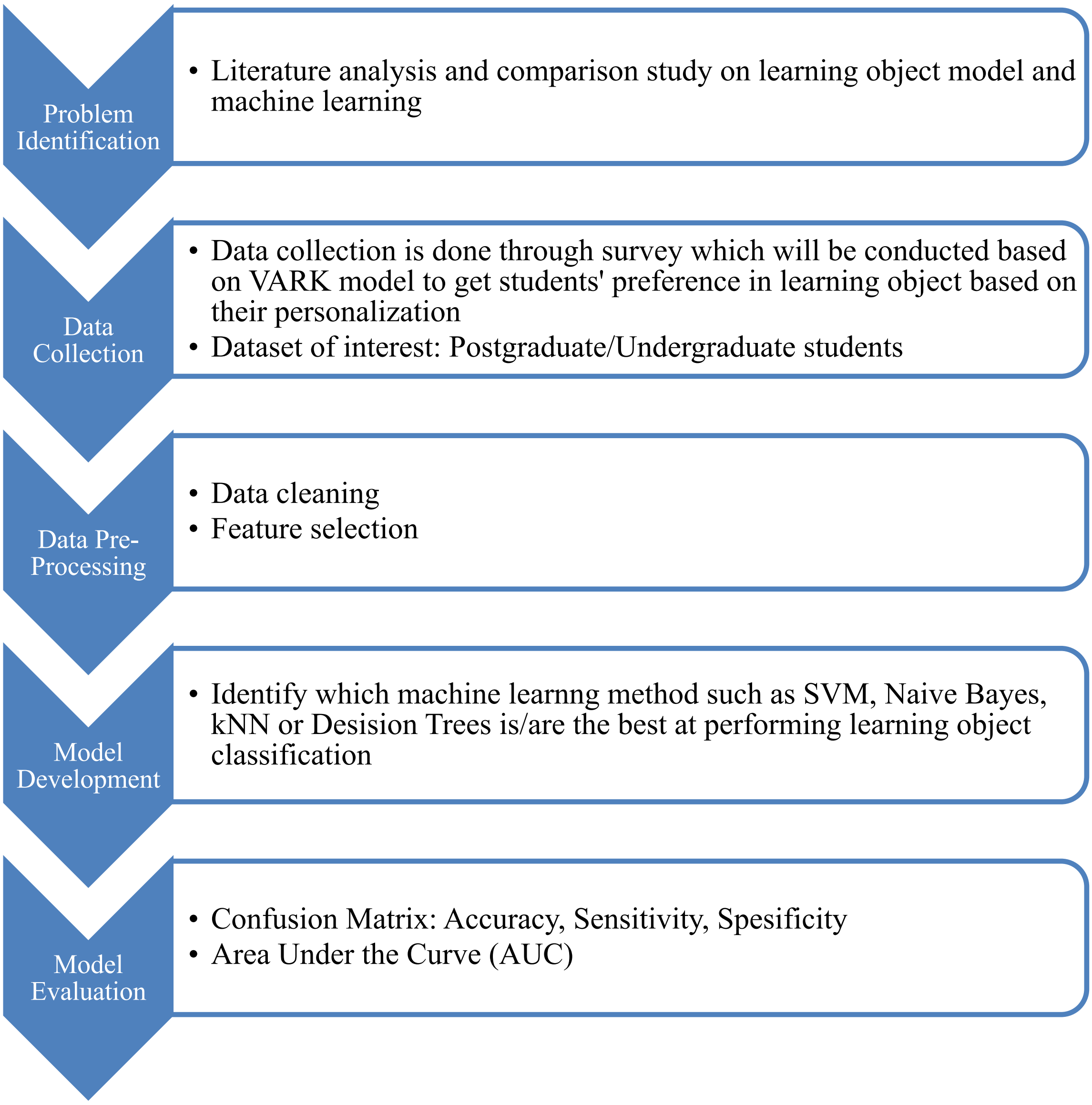


Figure 3.1: Process of Research

3.2 Problem Identification

Problem identification is mainly involving a literature review and preliminary study. The research problem is identified by reviewing previous work through literature analysis related to learning object, the concept of learning style, personalization, machine learning and information extracted by comparing all of the previous works.

3.3 Data Collection

As discussed in the previous section in Chapter 2, there are several learning models and each of the models offer various classification and representation with the types of learning. Each learner or student would have a significant attribute which is a unique learning object to provide personalized learning environment and to meet learning satisfaction. Hence, in this research project, the outcome will categorize the learners based on their learning style.

A survey questionnaire was conducted and distributed to the postgraduate and undergraduate students as the unit of analysis. For this research project, convenience sampling is applied, which resulted in the cumulative number of 966 respondents. Convenience sampling is a type of non-probability sampling technique where the sample is taken from a group of people that is contactable and reachable. A paper by Cresswell explained that in convenience sampling, a sample is selected from elements of a population who are willing to be studied and are easily accessible (Creswell, 2009). Question was framed based on the Index of Learning styles proposed by Felder-Silverman, VARK Model which defines four learning style such as visual, auditory, read/write and kinaesthetic. The respondents provide 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 gender, level of study, institutions, faculty, department, a field of study, country, household income, preferred learning mode, preferred social media platform, preferred communication platform and difficulties in online learning (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 about their learning style (three questions).
- Part 4 (Learning Style Test) – Testing students about learning style based on VARK learning style test (thirty questions).

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

3.4 Data Pre-Processing

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

3.4.1 Data Cleaning

Data cleaning was applied to the dataset in order to handle 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 answering questions thoughtfully. The following are the example of data cleaning done on the dataset.

- Removed respondents who only answer a portion of questions or leave blank or answer inappropriately.
- Removes respondents who do not meet the target criteria such as non-postgraduate or non-undergraduate students.
- Categorized as NA for the ambiguous response, doubtful results, no result especially in the open-ended question.
- Amended response with notation or symbol or special character such as “?”, “@”, and “/”.
- Z-score standardization was used to rescale the dataset for the purpose to improve the performance of the model.

3.4.2 Feature Selection

Originally, the dataset consists of 103 variables. The variables that were not contributing to the prediction, such as timestamp and identification, were removed appropriately. With the bag of words model, there will be massive features generated in the dataset. Hence, it is necessary to have the feature selection method in order to select the best features among these features. The motivation of feature selection is to achieve good accuracy result 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 research project, 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. Hence, 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.

3.5 Model Development

This project was based on the classification in supervised machine learning, which basically categorizes a set of data into classes to predict the class of the unlabeled 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 75% and 25%, respectively. Cross-validation (CV) is performed using the K-fold cross-validation technique. Ten-fold cross-validation is used to analyse the performance of all the algorithm. In ten-fold cross-validation, the training is set at random partition into ten folds, and each fold is used as a testing set at the same point. The final result is computed from the average of all performance matrixes to produce a single estimation. In this research project, there supervised machine learning models is used, which are Naïve Bayes, Support Vector Machine (SVM) and Decision Tree to classify the learning object.

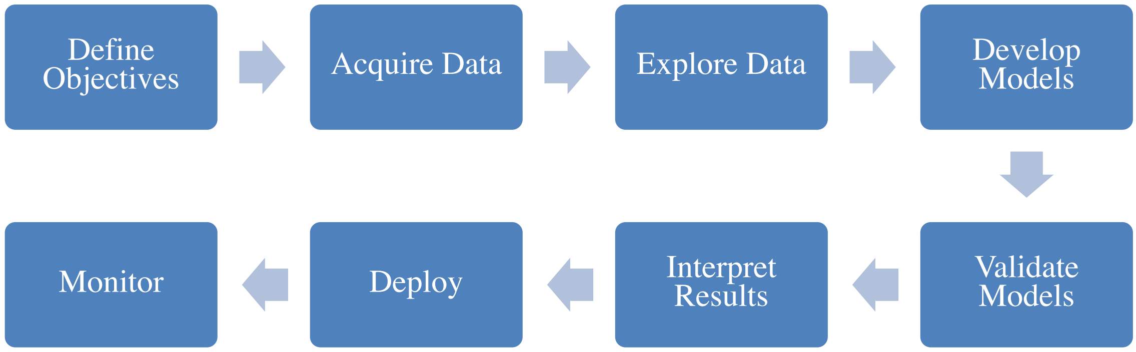


Figure 3.2 Model Development

3.5.1 Naïve Bayes

Naïve Bayes is commonly used in machine learning model for classification algorithm due to its simplicity and cost effective. It used 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 personalization. It will then classify the learning object into the class with the highest probability.

For the purpose to build Naïve Bayes, "e1071" package will be installed at beginning step. Similarly, with Support Vector Machines, we trained the Naïve Bayes model with the training data, setting the final learning object label as the target variable and then applied the model into the test data.

```
naive_bayes_model=naivebayes(final_learningobject ~., data=train_data)
test_pred<- predict(naive_bayes_model, newdata=test_data)
```

Figure 3.3: R Code To Build Naïve Bayes

3.5.2 Support Vector Machine (SVM)

SVM is s non-probabilistic classifier that can be used in classification. SVM separates data across a decision boundary determined by only a small subset of the data called a support vector. The boundary is called a hyperplane and will be optimized 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.

To build Support Vector Machines (SVM), “catools” package is installed in R Studio. The first step at this stage is to train the SVM with the training data, setting the final learning object label as the target variable and lastly, specified the method of SVM to be used. Then, at the end of the stage, the test data will be applied to the model.

```
svm_linear <- train(final_learningobject ~., data=train_data, method='svmlinear')
test_pred<- predict(svm_linear, newdata=test_data)
```

Figure 3.4: R Code To Build SVM

3.5.3 Decision Tree

Decision Tree is the most well-known approach to present classifiers. It consists of nodes which representing the features and the branches which called as decision node. There are a few basis for splitting the tree and growing the branches, such as the Gini index. Information gain is an impurity-based criterion that uses the entropy measure as the impurity measure. The feature which has the highest information gain will be the first node to be splitted. The pruning method is crucial in the decision tree as well in order to prevent the model to over-fitted. Pruning removes some sub-branches of the decision tree and rechecks the error rate. In the scenario if the error rate is no affected, the sub-branches will be removed accordingly.

In order to build the third model namely Decision Tree by installing "rpart" package. We trained the decision tree model with the training data, setting the final learning object label as the target variable and then applied the model into the test data.

```
tree<- rpart(final_learningobject~., data=train_data)
test_pred<- predict(tree,type="class",newdata=test_data)
```

Figure 3.5: R Code To Build Decision Tree

3.6 Model Evaluation

Once all the models are trained, each and every model will then be being tested with the test data. The performance of the model will be judged and measured from the accuracy, sensitivity and specificity of the prediction, which can be computed from the confusion matrix below.

Table 3.1: Confusion Matrix

	Actual Positive	**Actual Negative**
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

Accuracy is the percentage of the data which are predicted correctly.

$$Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Data}$$

Meanwhile, specificity can be defined as how accurate the model in detecting the actual positive data out from all of its predicted positive data.

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

Lastly, sensitivity is the proportion of actual positive data which is identified correctly.

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

Eventually, the area under the curve (AUC) is derived from a summary measure of accuracy from ROC curves. In the evaluation of this research project, the model which has the higher accuracy will be identified as the best model among others.

3.7 Summary

This chapter has addressed the process involved in developing the model, including an introduction to the framework and the experiments. To conclude, this chapter underlined some related and 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 the learning object.

CHAPTER 4: FINDINGS AND DISCUSSIONS

4.1 Introduction

This chapter is divided into two main sections, namely findings and discussion. The result part presents the outcome attained during model development to predict learning object using machine learning approaches for classification. Thereafter, the discussion section elaborates on the data description, explanation on data pre-processing and evaluation of model performance.

4.2 Findings

4.2.1 Part 1: Data Exploratory Analysis

A survey is conducted to study online learning and students' learning preference of undergraduate and postgraduate students. The details of the survey are described briefly in Section 3.2.

The following table shows the distribution of the respondents in terms of their demographic profile.

Table 4.1: Demographics Profile

Demographic Profile (n=966)		**Responses (N)**	**Responses (%)**
Gender	Female	649	67.2%
	Male	317	32.8%
Level of Study	Certificate/Diploma	132	13.7%
	Undergraduate	818	84.7%
	Postgraduate	5	0.5%
	Master	8	0.8%
	PhD	3	0.3%
Field of Study	Computer Science/Information Technology	145	15.0%
	Architecture and Building	101	10.5%
	Engineering	92	9.5%
	Economy	66	6.8%
	Education	57	5.9%

	Business and administration	49	5.1%
	Accounting	46	4.8%
	Sports	41	4.2%
	Arts	38	3.9%
	Pharmacy	16	1.7%
	Others	315	32.6%
Institution	UM	549	56.8%
	UITM	144	14.9%
	KKTM	89	9.2%
	NA	83	8.6%
	IIMAT	22	2.3%
	UIA	22	2.3%
	MAHSA	21	2.2%
	KOLEJ KOMUNITI	7	0.7%
	USM	7	0.7%
	MSU	6	0.6%
	Others	16	1.7%
Country	Malaysia	951	98.4%
	Bangladesh	5	0.5%
	China	3	0.3%
	India	3	0.3%
	Somalia	1	0.1%
	Sudan	1	0.1%
	Egypt	1	0.1%
	Indonesia	1	0.1%
Household Income	Less than RM 4,849	587	60.8%
	RM 4,850 – RM10,959	290	30.0%
	More than RM10,960	89	9.2%

- Preferred Learning Mode

This question is meant to address students preferred learning mode. Figure 4.1 reflected the results of students' preference for their learning mode.

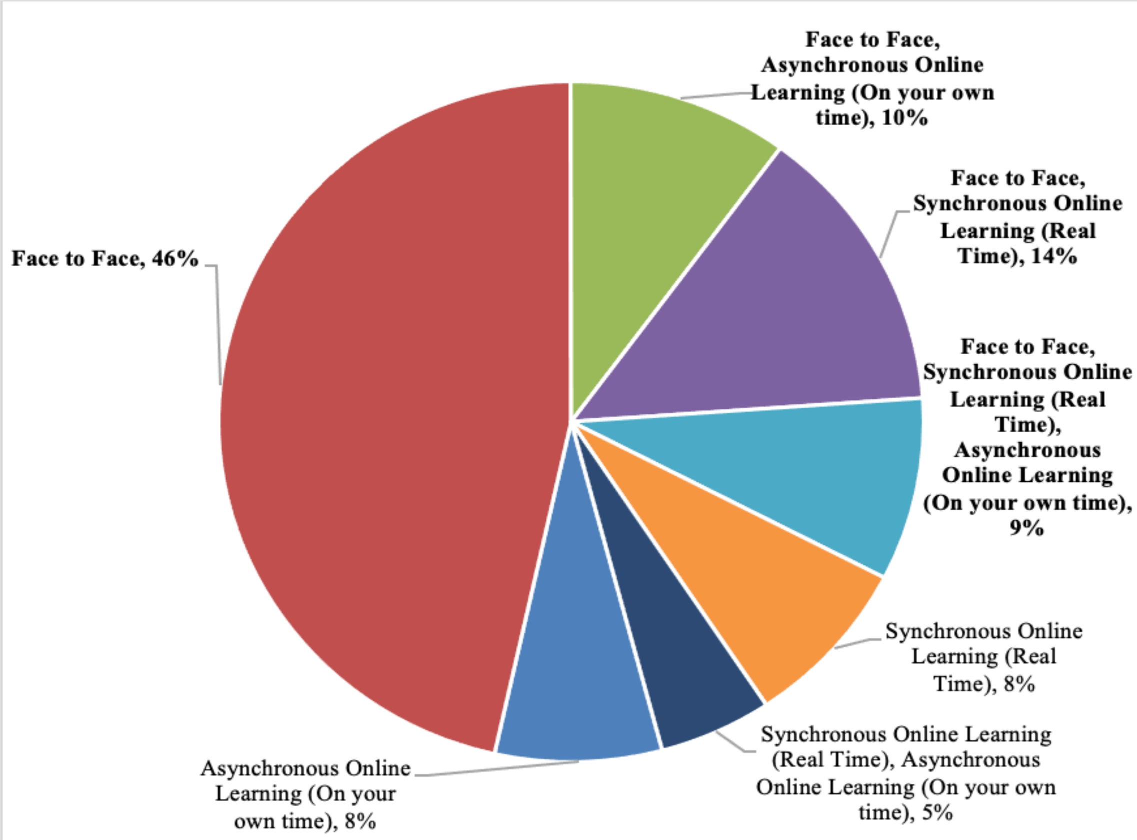
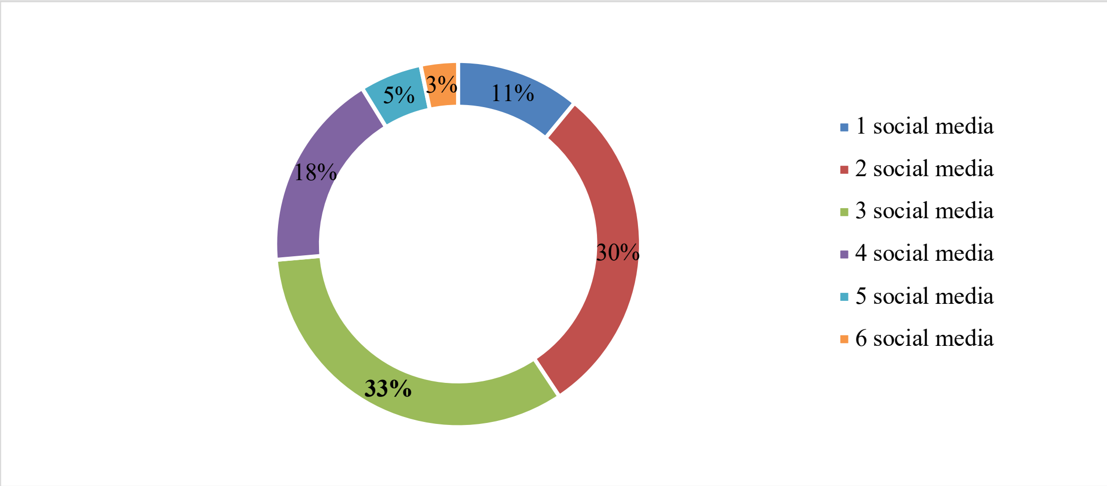


Figure 4.1: Preferred Learning Mode

- Preferred Social Media Platform

In the era of modern and digital technology, there are various forms of social media platform exist which most of the students will be having at least one. This platform is benefitting the students mostly in engagement and communication. The following figures indicate the number of accounts a student has and their favourite social media platform.



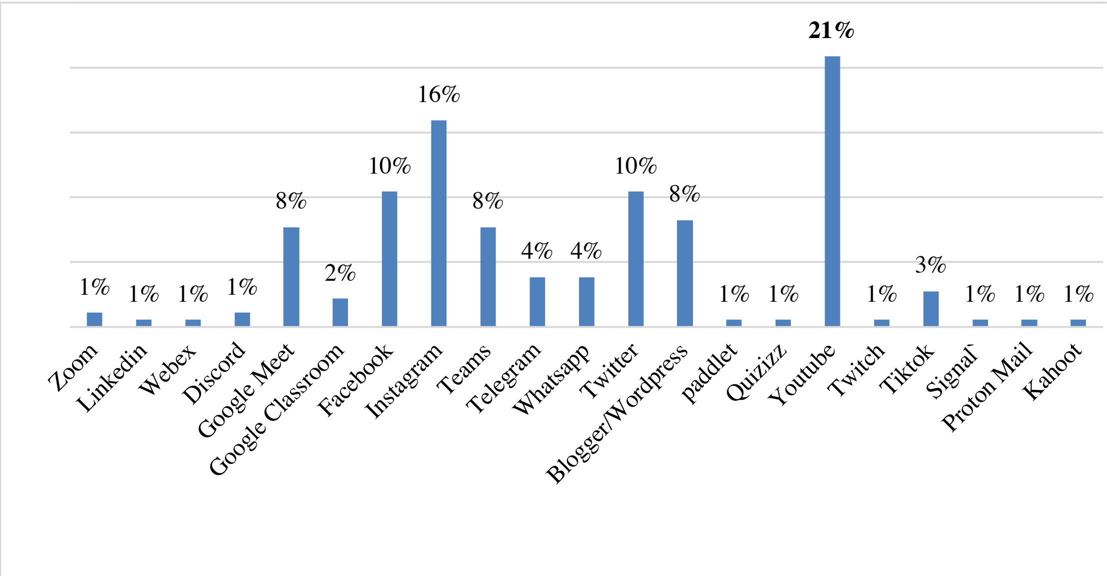


Figure 4.2: Preferred Social Media Platform

- Preferred Communication Platform

In this question, the objective is to identify the student's preference in communication during their learning process. With the variety of channel available from Whatsapp, Telegram and many others, students tend to use those channel in their effective communication for learning.

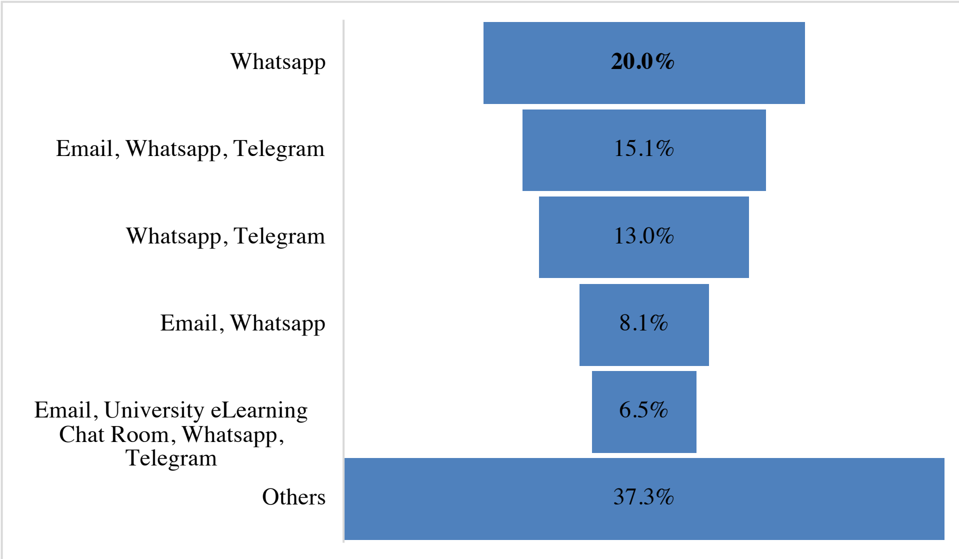


Figure 4.3: Preferred Communication Platform

- Difficulties in Online Learning

This question demonstrates difficulties faced by students in online learning. There is various of the answer provided as the respondents were allowed to choose more than one available options. Word cloud illustrated below shows the common difficulties faced by students in online learning.

Figure 4.4: Word Cloud Of Difficulties Of Learning

- Learning Object

The study reveals the preference of students to use learning object and online instructional strategies in their study. The answer is a scale from Not At All until Very Much. Learning object consists of various options from a slide presentation to intelligent computer-aided instruction systems. The table below shows the preferred learning object.

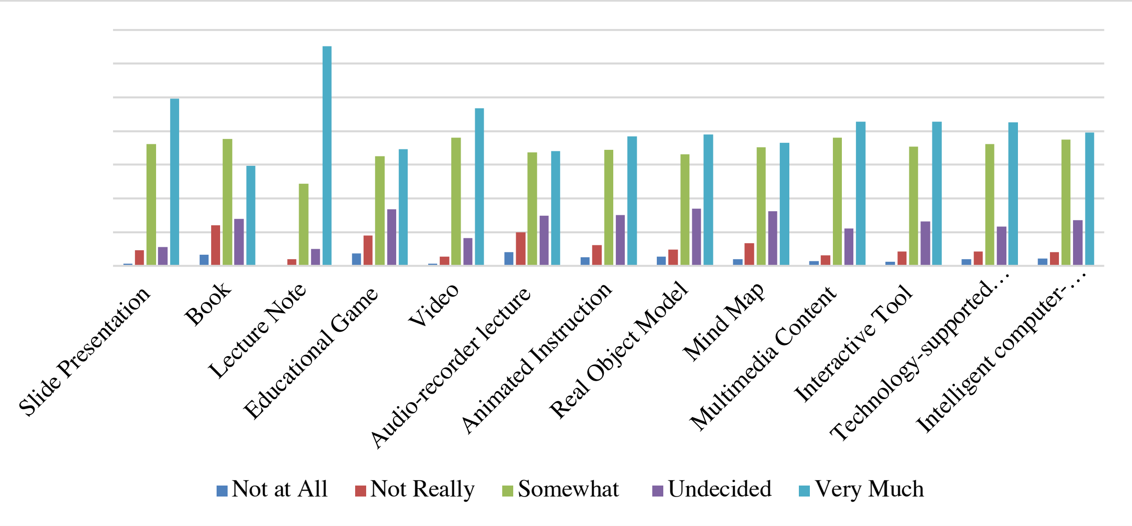


Figure 4.5: Learning Object

- Learning Style Awareness

For the purpose of this survey, learning style is defined as the predisposition or preference of an in a particular way or a combination of ways individual to perceive and process information. This section consists of three questions to indicate student's awareness whether they are familiar, aware and knowing the importance of learning style to improve the learning ability. The figures illustrated below indicates the findings.

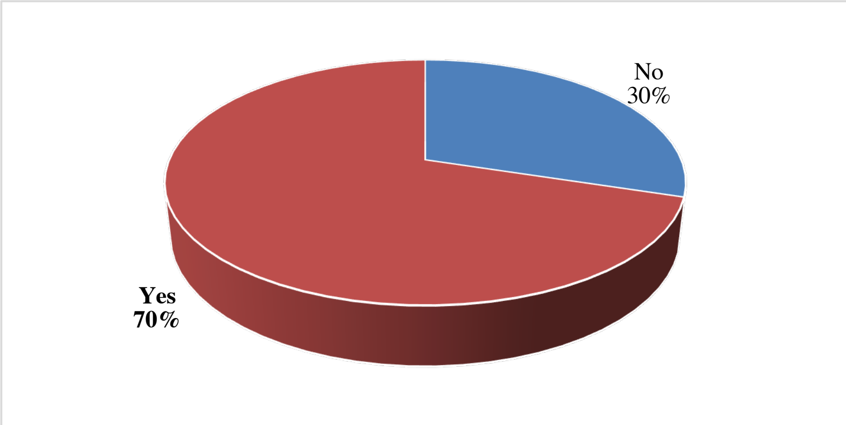


Figure 4.6: Learning Style Familiarity

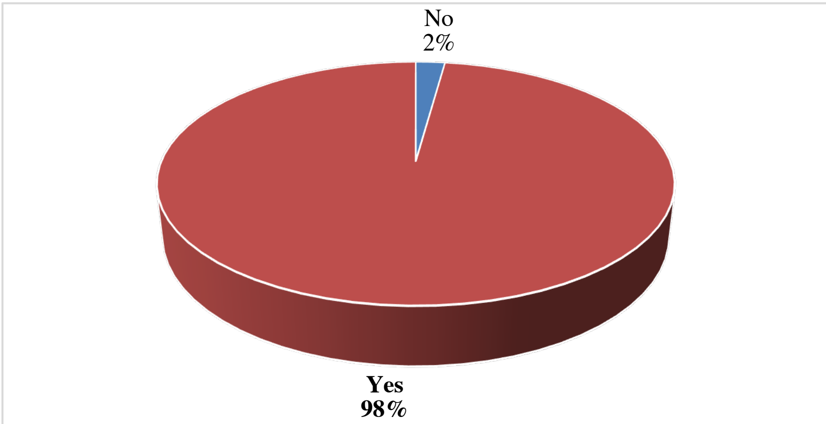


Figure 4.7: Learning Style Awareness

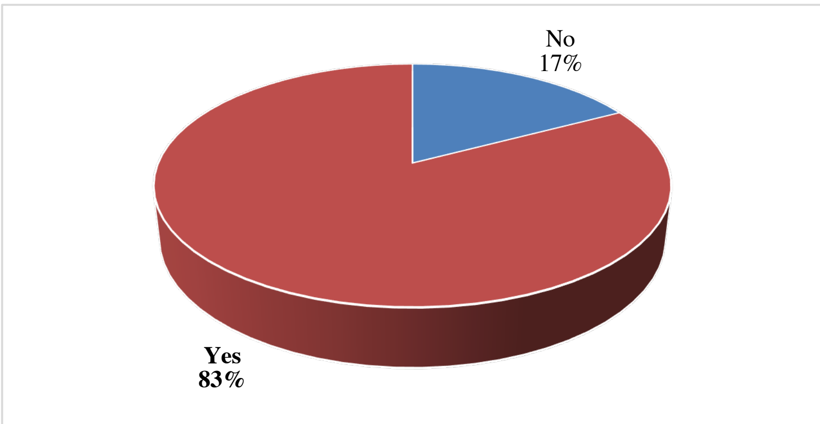


Figure 4.8: Degree of Knowing The Importance of Learning Style

This behavior is supported by the following comments at the end of the questionnaire.

"Different method for different learning style. Teacher should be trained with exposure of different learning style ."

"Feels like I've rediscovered myself"

4.2.2 Part 2: Data Modelling

In this section, we will split the data into training and test data set. We used the training data to develop the three models. At the end of this section, the performance of all models is evaluated and compared accordingly to indicate the best model.

4.2.2.1 Data Partition

The dataset is split into 75% of training data and 25% of test data. The data is selected in random into train and test data. Train data is used to build the predictive model and test data is used to test the accuracy of the model.

```
set.seed (123)
sample = sample.split(final.df,splitratio=0.75)
train_data=subset(final.df,sample==TRUE)
test_data=subset(final.df,sample==FALSE)
```

Figure 4.9: R Code For Data Partition

4.2.2.2 Model Building

In this research project, as discussed in the previous section, three models are built to predict the learning object. They are namely Support Vector Machine (SVM), Naïve Bayes and Decision Trees.

In order to see the number of predicted and actual learning object of the test data, confusion matrix is built for each model. By this information retrieved, the models are evaluated based on their accuracy, specificity and sensitivity.

```
confusionMatrix(test_pred, test_data$final_learningobject)
```

Figure 4.10: R Code To Build Confusion Matrix

From the matrix result, we computed the accuracy, specificity and sensitivity rate of each model.

Table 4.2: Result of Accuracy, Specificity and Sensitivity

Model	**Accuracy (%)**	**Sensitivity (%)**	**Specificity (%)**
SVM	70.43	67.89	91.72
Naïve Bayes	43.71	27.92	80.42
Decision Tree	54.36	50.75	90.88

4.3 Discussion

4.3.1 Part 1: Data Exploratory Analysis

A total of 966 respondents were identified to respond to the questionnaire. 67.2% of the respondents are female, followed by male at 32.8%. The respondents are grouped into certificate/diploma students (13.7%), undergraduate students (84.7%), postgraduate students (0.5%), master students (0.8%) and PhD (0.3%). Most of the respondents were from Computer Science/Information Technology course with 15.0% and followed by Architecture and Building, Engineering and others with the percentage of 10.5%, 9.5% and 65.0% respectively. The survey has been distributed across many institutions and the highest respondents were from the University of Malaya which represents half of the respondents, 56.8%. The majority of respondents were Malaysians, which constitutes 98.4%, and the remaining is foreign students from Bangladesh, India, Somalia, Egypt, Sudan, Indonesia and China. In terms of household income, 60.8% fall under less than the RM4,849 income bracket and only 9.2%% 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, even for multiple answers, the mode is still within the choice among the respondents. The majority of the respondents were having three social media accounts and Youtube is the most popular social media platform among them, followed by Instagram, which contributes 21% and 16%, respectively. Whatsapp has the highest percentage compared to the other channel, about 20%. 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 technical issue, followed by commitment and focus. The findings of the preferred learning object revealed that most of the students opt for lecture note as their learning object. Apart from that findings, the majority of the students answer Live Lecture for the technical or hands-on subject in terms of online instructional or assessment preference. Most of the students agreed and 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

In the previous section, we have seen how the three models could predict the learning object based on the accuracy rate across three models. However, for the second model, which is trained using Naïve Bayes, it has the lowest accuracy compared to Support Vector Machine and Decision Tree with the rate stood at 43.71%, albeit the training time took run this model is the shortest among the three models used in this research project (Ouafae et al., 2019). As mentioned in the literature, Naïve Bayes works well if the training set is small, since this research is having quite massive number of rows and columns, hence this might be the contributed reasons Naïve Bayes cannot perform well.

Support Vector Machine has the highest rate of accuracy, which is 70.43% and. The higher the validation matrics, the good is the classifier. As observed, all the validation metrics, including accuracy, sensitivity and specificity, have high values for Support Vector Machines; hence we can conclude that the classifier used in the approach has well performed. For this stated reason, most learning object classifiers have move on to using Machine Learning approach because of its high accuracy and the ability to classify the learning object based on the learning style model. This means that when an educator wants to find the synonym learning object for example, a Support Vector Machine model is able to understand the context and provide similar meanings for the correspond learning style.

For Decision Tree, its overall accuracy is lower than Support Vector Machine. Overall, Support Vector Machine is the best model in this research due to its accuracy, sensitivity and specificity rate. The result from this research in beyond the expectation based on literature due to the dimension of data which cause Naïve Bayes not able to perform as per previous paper.

4.4 Summary

In this chapter, we started with exploratory data analysis from the result of the survey questionnaire. We have presented the findings into visualization and with some insight in the subsequent discussion. We then split the data into training and testing data to run the model based on three machine learning approach, namely Support Vector Machine, Naïve Bayes and Decision Tree. The model is then evaluated in the discussion to identify the best model for the classification of learning object based on personalization. As a conclusion, Support Vector Machine is found to be the best model in this study.

CHAPTER 5: CONCLUSION

5.1 Introduction

This research project focuses on utilizing the machine learning approach and its application in predicting learning object. This chapter will make a conclusion on the research project by revisiting the research objectives and answering the research questions. Discussion on limitation also will be covered as well as the future work to be carried out.

5.2 Summary of Findings

RO 1: To develop a classification model of learning object based on personalization.

RQ 1: How should the classification model of learning object based on personalization be developed?

We achieved the first objective of the research project, which is to develop a classification model of learning object based on personalization. The classification model is developed by collecting the data from survey questionnaire on students' feedback and performed pre-processing method to clean the data. By doing data cleaning, we able to get a better quality of data and generate features from the data. With this feature, we successfully classify the learning objects by personalization by applying machine learning models namely Support Vector Machine, Naïve Bayes and Decision Tree.

RO 2: To evaluate the model performance of learning object based on personalization.

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

We had also successfully achieved the second objective to evaluate the model performance of the learning object by finding the best machine learning model in this research. This research found Support Vector Machine to be the best classifier model. This can be explained by SVM when the accuracy rate is the highest among the other model.

5.3 Limitation

There are several limitations to our approaches used in this study. Firstly, the research only compared three machine learning models as there are few more machine learning models which can be used and to be compared with their performance. For the purpose to improve learning capabilities, the study on learning object and learning style should be extensively be done across various machine learning algorithm.

In terms of data availability, the dataset used is relatively low where the total target respondents are expected to reach at least two thousand. Furthermore, we had analysed the learning objects limited to university students only which include undergraduate and postgraduate students without primary or secondary students' penetration which should be covered holistically in education system. We did not include primary and secondary students to control the size and focus on the target criteria.

5.4 Future Work

To study more about learning object among students in education system, students' feedback is highly recommended to acquire and strengthen the information. In the future study, we can maximize the learners or student's definition among primary and secondary students, which eventually can bring benefits to the whole education system holistically. Besides, this might give us some new insights as the group of people who is responding to the questionnaire are different.

Apart from that, we can try other new predictive models as well in the research. As this research had proven the Support Vector Machine is the best model to classify the learning object, we can use this model as a benchmark to compare with the other models such as Neural Network, which had not been studied in this research project.

5.5 Conclusion

This study intends to discover the potential of using machine learning to complement the conventional method in classifying learning object in the education system. The performance of the classifiers was evaluated and compared accordingly. The results strongly skewed to Support Vector Machine to be the promising approach in predicting learning object based on personalization. To the best extend, this is the first study in cooperating student feedback in the research, which assist in closing the gap with previous work done. The approach of this research project will provide a significant impact and mutual benefit between learners and educators.

REFERENCES

Dey, A. (2016). Machine learning algorithms: a review. International Journal of Computer Science and Information Technologies, vol. 7, no. 3, pp. 1174–1179.

John Martin, A., Maria Dominic, M. & Sagayaraj Francis, F. (2021). Learners Classification for Personalized Learning Experience in e-Learning Systems International Journal of Advanced Computer Science and Applications(IJACSA), 12(4) http://dx.doi.org/10.14569/IJACSA.2021.0120485.

Janikow, C. Z. (1998). Fuzzy decision trees: issues and methods," IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 28, no. 1, pp. 1–14.

Cassidy, S. (2004) Learning styles: An overview of theories, models, and measures. Educational Psychology, 24 (4), 419-441.

El Aissaoui, O., El Alami El Madani, Y., Oughdir, L. & El Allioui, Y. (2019). Combining supervised and unsupervised machine learning algorithms to predict the learners' learning styles. Second International Conference on Intelligent Computing in Data Sciences, (pp. 87-96).

Franzoni, A. & Assar, S. (2009). Student learning styles adaptation method based on teaching strategies and electronic media. Educational Technology & Society, 12(4), 15–29. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4561832.

Stein, G, Chen, B, Wu, A. S. & Hua, K.A. (2005). Decision tree classifier for network intrusion detection with GA-based feature selection," in Proceedings of the 43rd annual Southeast regional conference-Volume 2, 2005, pp. 136–141.

Haq, A. & Chand, S. (2012). Pattern of Facebook usage and its Impact on Academic Performance of University Students : A Gender Based Comparison*,* 34(2), 19–28.

James, W. B. & Gardner, D. L. (1995) Learning Styles: Implications for Distance Learning. New Directions for Adult and Continuing Education no. 67, 19-32. (1) (PDF) Learning Styles and Their Relation to Teaching Styles. Retrieved from: https://www.researchgate.net/publication/275567766_Learning_Styles_and_Their_Relation_to_Teaching_Styles.

Kim, J., Jung, Y., Lim, Y. & Kim, M. (2009). An E-Learning Framework Supporting Personalization and Collaboration in ACM 3rd International Conference on Ubiquitous Information Management and Communication (ICUIMC).

Kaya, G. (2011). A Learner Model for Learning Object Based Personalized Learning Enviro. Retrieved from https://link.springer.com/chapter/10.1007/978-3-642-24731-6_35?error=cookies_not_supported&code=44a448d0-b0ad-45e1-b1fd-ca155d4b0121.

Kolb, D.A. (1984). Experiential learning: experience as the source of learning and development (11th ed). New Jersey-Hall.

Martin, F. (2020). A synthesis of systematic review research on emerging learning environments and technologies. Educational Technology Research and Development. https://link.springer.com/article/10.1007/s11423-020-09812-2?error=cookies_not_supported&code=c9417b30-a251-4f29-898f-9c46ae1e3344.

Martinez, M. (2000). Designing Learning Objects to Mass Customize and Personalize Learning. In: Wiley, D.A. (ed.) Instructional Use of Learning Objects: Online Version.

Mestre, L. S. (2010). Matching Up Learning Styles with Learning Objects: What's Effective? *Journal of Library Administration,* 50(7-8), 808–829. doi:10.1080/01930826.2010.488975.

Normadhi, N., Shuib, N., Md Nasir, H., Bimba, A., Idris, N., & Balakrishnan, V. (2019). Identification of personal traits in adaptive learning environment: Systematic literature review. Computers & Education , 168-190.

Reid, J.M. (1995) Learning Styles in the ESL/EFL Classroom. Boston: Heinle & Heinle Publishers.

Gamalel-din, S.A. (2012). An Intelligent eTutor-Student Adaptive Interaction Framework," in ACM INTERACCION '12 Proceedings of the 13th International Conference on Interacción Persona- Ordenador.

Ghallabi, S., Essalmi, F., Jemni, M. & Kinshuk (2013). Toward the reuse of E-Learning personalization systems in IEEE 2013 Fourth International Conference on Information and Communication Technology and Accessibility (ICTA).

Sampson, D., Karagiannidis, C. & Kinshuk. (2002). Personalized learning: Educational, technological and standardization perspective. *Interactive Educational Multimedia 4*, 24–39.

Shuib, N. (2013). Information seeking tool based on learning style. UM Students' Repository. http://studentsrepo.um.edu.my/5605/.

Smith, R. (2004). Guidelines for Authors of Learning Objects (pp. 1–32). *NMC: The New Media Consortium.*

Sun, Zehang, George Bebis, and Ronald Miller (2004). Object detection using feature subset selection." Pattern recognition 37.

Truong, H. (2016). Integrating learning styles and adaptivee learning system: Current developments, problems and opportunities. Computers in Human Behavior , 1185-1193.

Williamson, MF. & Watson, R. (2006) Learning Styles Research: Understanding How Teaching Should Be Impacted by the Way Learners Learn .Christian Education Journal, 3(1), 27-43.

Zhang, Tianzhu, Si Liu, Changsheng Xu, and Hanqing Lu. (2013). Mining semantic context information for intelligent video surveillance of traffic scenes." Industrial Informatics, IEEE Transactions on 9.