A DYNAMIC ELEARNING PREDICTION MODEL BASED ON INCOMPLETE ACTIVITIES OF ELEARNING SYSTEM

SONGSAKDA CHAYANUKRO

DOCTOR OF PHILOSOPHY
UNIVERSITI UTARA MALAYSIA
2020

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

SONGSAKDA CHAYANUKRO

calon untuk Ijazah **PhD**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"A DYNAMIC ELEARNING PREDICTION MODEL BASED ON INCOMPLETE ACTIVITIES OF ELEARNING SYSTEM"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***13 Januari 2019.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
January 13, 2019.

Pengerusi Viva: *(Chairman for VIVA)*	**Assoc. Prof. Dr. Fauziah Baharom**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Dr. Mohd Nasir Ismail**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Assoc. Prof. Dr. Mazida Ahmad**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Massudi Mahmuddin**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Dr. Husniza Husni**	Tandatangan *(Signature)*

Tarikh:
(Date) ***January 13, 2019***

Permission to Use

Abstrak

Pada masa kini, penggunaan e-Pembelajaran adalah pelbagai kerana aktiviti e-Pembelajaran yang digunakan dalam pengajaran dan pembelajaran berbeza bergantung kepada pendidik. Pemilihan aktiviti dalam penggunaan e-Pembelajaran yang berbeza mempengaruhi ramalan hasil pembelajaran. Walau bagaimanapun, kebanyakan model ramalan hasil e-Pembelajaran masih tidak stabil dan tidak dapat digunakan dalam pelbagai situasi kerana penggunaan e-Pembelajaran dianggap sangat dinamik. Oleh yang demikian, objektif kajian ini adalah: a) untuk menganalisis aktiviti e-Pembelajaran yang mempengaruhi hasil pembelajaran; b) untuk membina model ramalan hasil pembelajaran bagi penggunaan e-Pembelajaran; c) untuk mensintesis model ramalan e-Pembelajaran dinamik berdasarkan aktiviti tidak lengkap sistem e-Pembelajaran; dan d) untuk menilai model ramalan e-Pembelajaran dinamik berdasarkan kelebihan, ketepatan, dan keberkesanannya. Kajian ini dijalankan dengan tujuh langkah: kajian awal; pengumpulan data; pemprosesan data; analisis aktiviti e-Pembelajaran; pembinaan model ramalan hasil pembelajaran; pensintesisan model ramalan e- Pembelajaran; dan penilaian model. Enam algoritma perlombongan data telah digunakan dalam penilaian model. Hasil kajian mendapati tujuh kumpulan penting aktiviti e-Pembelajaran yang dapat meramal hasil pembelajaran dengan ketepatan melebihi 75%. Daripada tujuh kumpulan penting tersebut, dua kumpulan aktiviti mempunyai nilai Receiver Operating Characteristic melebihi 0.5. Oleh itu, kajian ini menunjukkan bahawa penggunaan data daripada aktiviti tidak lengkap sistem e-Pembelajaran menyediakan cara yang sesuai bagi hasil pembelajaran yang boleh diramal. Model ramalan ini menyumbang kepada bilangan kelas dan set data yang optimum di mana dua kelas menerima nisbah ketepatan tertinggi. Secara praktikal, hasil kajian ini boleh membantu menambah baik pengurusan dan mengurangkan kos pendidikan.

Kata kunci: e-Pembelajaran, aktiviti e-Pembelajaran tidak lengkap, model ramalan e-Pembelajaran, ramalan hasil pembelajaran, Sistem Pengurusan Pembelajaran.

Abstract

At present, eLearning usage is diverse because the eLearning activities used in teaching and learning differ depending on educators. The selection of activities in different eLearning usage affect the prediction of learning outcomes. However, most eLearning outcome prediction models are still unstable and inapplicable in many situations as the eLearning usage is considered to be highly dynamic. Therefore, the objectives of this study are: a) to analyze the eLearning activities that affect learning outcome; b) to construct a learning outcome prediction model for eLearning usage; c) to synthesize a dynamic eLearning prediction model based on incomplete activities of eLearning systems; and d) to evaluate the dynamic eLearning prediction model based on advantage, accuracy, and effectiveness. This study was conducted through seven steps: initial study; data collection; data pre-processing; eLearning activity analysis; learning outcome prediction model construction; eLearning prediction model synthesizing; and model evaluation. Six data mining algorithms were used in evaluating the model. The results found seven significant groups of eLearning activities that could predict the learning outcome with more than 75% accuracy. Of the seven significant groups, two groups of activities have Receiver Operating Characteristic values greater than 0.5. Hence, this study demonstrates that using data from incomplete activities of eLearning systems provides an appropriate means for predictable learning outcomes. The prediction model contributes to an optimal number of classes and data set where two classes received the highest accuracy ratio. Practically, the results of this study may assist towards improving management and reducing educational costs.

Keywords: eLearning, eLearning incomplete activities, eLearning prediction model, learning outcome prediction, Learning Management System.

Acknowledgement

Firstly, I would like to express my deepest gratitude to my dear life-coach, my beloved mom and my family who made me choose the Ph.D. journey.

A very special thank goes out to my supervisor and co-supervisor, Dr. Massudi bin Mahmuddin and Dr. Husniza binti Husni, who always opened their office’s door for me whenever I ran into a trouble and had a question on my research. Their patience, motivation, enthusiasm and immense knowledge handed over to me in all time of my reserach journey. I couldn’t have imagined having better advisors and mentors for my Ph.D. study.

For me, this long journey also offered a great experience from my new friends from many different countries, races and religions, but in fact, good friendships are borderless. I am also so thankful to my close friends who have shared a great time with me at this university and given me so much love and help, especially Miss Saowatarn Samanit, who encouraged and directed me.

My special and heartily thanks to Mr. Charnpich Attajak, my first roommate who hasn’t got a chance to get his Ph.D. and never ever say the word “congrats” to me on this success as he passed away in a car accident.

Additionally, I am also grateful to the executives and my colleagues from Suan Dusit University who have given me an opportunity to further my Ph.D. and provide funding for this work.

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

ADL	Advanced Distributed Learning
API	Application Programming Interface
CBT	Computer Based Training
CMS	Course Management System
CPE	Continuous Professional Education
DBMS	Database Management System
DM	Data Mining
GA	Genetic Algorithm
ITS	Intelligence Tutorial System
LMS	Learning Management System
LOPM	Learning Outcome Prediction Model
ML	Machine Learning
MOODLE	Modular Object-Oriented Dynamic Learning Environment
ROC	Receiver Operating Characteristic
ROI	Return on Investment
RTE	Run-Time Environment
SCORM	Sharable Content Object Reference Model
SVM	Support Vector Machine
TTA	Timed Transition Automaton
VLE	Virtual Learning Environment
WBT	Web based training
WEKA	Waikato Environment for Knowledge Analysis
WUM	Web Usage Mining

CHAPTER ONE
INTRODUCTION

1.1 Introduction

As the information technology age has progressed and changed, the learning environment has developed into an eLearning system and has facilitated the learning process. The impact of globalization has made the university more competitive in offering higher quality education and flexibility. Therefore, creating a clear vision and framework for implementing strategies for educational innovation and eLearning is essential (The & Usagawa, 2017). Presently, eLearning concept is viral among higher education institutions (Dai & Zhang, 2008). The lecture materials are allowed by an eLearning system for learners to learn and experience through the network (Min, 2005).

Nowadays, eLearning is remarkably developed. The system has tremendous changing to facilitate classroom for more efficient outcome comparing to a traditional class. This modern teaching process, in fact, applies advanced communication technologies which do not only fulfill the need of educational institutions, but it can also build the efficient communication in the digital classroom. eLearning is a widely used technology, and it can assist in providing guidance and analyzing the effects of classroom management. Besides, the use of eLearning in teaching management helps to understand the impact of eLearning. It is the ubiquitous system.

Currently, the Learning Management System (LMS) was developed to meet the requirement of the above developing virtual classroom. Therefore, higher learning institutions are finding suitable LMS for eLearning system. There are various ways to select LMS tools in the virtual class (Graf & List, 2005); so, open source LMS is now the most significant tools which applies to the learning environment. The study of open source LMS in eLearning has found Moodle LMS as the most outstanding eLearning system (Graf & List, 2005). Moreover, there are also many useful features in LMSs Moodle that strive for building on quality of education and give assistance to select the necessary tools for eLearning.

According to the rapid development of eLearning to broaden the use of open source software provides the evolution of learning tools and quality of education. Furthermore, the most significant advantage of eLearning is that it is less cost consuming than traditional learning environments (Aydin & Tirkes, 2010). The Learning Management System (LMS) has been developed to introduce activities to refine individual learning objectives. All learners have the opportunity to create different activities but from the same learning objectives. Course instructors can create various learning activities and contents for one learning objective based on the field of interest. Therefore, it is possible to convey the concept of learning objectives from different types of activities. This difference is created by analyzing students' dynamic behavior in the virtual learning environment (VLE) (Gunathilaka, Fernando, & Pasqual, 2017).

Moodle is an eLearning designed for the university that aims for a virtual campus. The system majorly comprises of a Learning Management System (LMS). The result is that the eLearning system has an impact on the interactions between learners and teachers. They can set the specific values that benefit the learner in participating in certain activities. At the same time, some higher education institutions have developed higher eLearning platforms to meet the university's concept of virtualization (Andone, Ternauciuc, & Vasiu, 2017; Huertas & Navarro, 2017).

Concerning the impact of eLearning, it has been studied and found that eLearning can assist either student or teacher to stimulate students' knowledge and manage learning time. And more importantly, the learning experience through eLearning allows the user to interact with the social community (Wardaya & Pradipto, 2017). In India, it showed that the use of eLearning could be grouped differently. The grouping does not depend on the social status and economic status anymore. It was evident that after the application of eLearning, students have changed their teaching and learning behaviors. It is advantageous that taking advantage of eLearning methods that are more distinctive, the gap in educational services has decreased (Gulati, Batra, Khurana, & Tripathi, 2017).

The development of advanced eLearning systems has introduced a smart English learning system that works on web pages. This system utilizes data mining techniques to group student learning patterns based on learning style. The ultimate goal of the system is to define the best teaching style for each learner, which can be used anywhere on the web. There are also facilities like videos for learning and quizzes. This system allows the instructor and learner to follow the best learning path. The results showed that the student achievement model was at the level of 87.4% (Tashtoush, Al-Soud, Fraihat, Al-Sarayrah, & Alsmirat, 2017).

In higher education institutions, eLearning is a highly innovative educational innovation. Higher education institutions have adopted this eLearning data to enhance the value of education. Advantages of eLearning are as the following reasons (Homiakova, Arras, & Kozík, 2017):

- Electronic learning is not so expensive since there is no requirement of traditional classroom equipment.
- Learners can study whenever and wherever they want, regardless of geographical location.
- eLearning is an integral part of the long-term strategy of higher education institutions.
- In the view of eLearning learners, it increases the opportunity to interact with other learners and instructor, and it can access various multimedia resources from experts all over the world.
- eLearning is effective in evaluating distant learners through its tools.

Following the system that supports eLearning usage variables, several relevant research findings prove the above discussion (Aydin & Tirkes, 2010; Graf & List, 2005). In fact, the study of Moodle eLearning system with an open source data structure found many activities relevant to learning usage such as student activities. Consequently, these student activities are the source of the web log that can apply to construct the usage model by data mining techniques (Romero, Ventura, & Garcia, 2008).

eLearning system platforms share some similar functions such as the educational administration function, management of education resources, management of curriculum, management of collaboration, management of evaluation and aids tools (Jing, Hailong, & Jun, 2009). To improve the efficiency of eLearning systems, the study of learners 'learning styles is a critical part in enhancing learners' performance. Understanding the learners' immense learning styles will help them to access the learning outcomes of their learners through eLearning.

In general, learning outcomes can be predicted according to learners' characteristics through various patterns of learning styles. Moreover, a learner's behavior and cognitive skills can define learner's learning styles. eLearning is primarily intended to provide a communication platform for learning management. The development of an eLearning system called the intelligence tutorial system (ITS) has been developed to guide the needs of learners and instructors. The results of this system can better determine the level of knowledge and the concept of learners (Deena & Raja, 2017).

Some researches intend to generate a functioning eLearning system basis comprising students' requirements and suggest the system workflow to and from the functional modules of the system. According to Chun et al. (2010), it is found that instructors must focus more on learner's enthusiasm and their most popular digital learning tools (Chun Xia, Hui Bao, Chang Yi, & Yue Xing, 2010). According to the above arguments, exciting critics have been raised in the angle of creating a proper model up to learner needs. Analysis of web log by data mining techniques for demonstrating or discovering new knowledge hidden in an extensive database is a challenging task. Several researchers mentioned earlier have been obtained the prediction model from rather wholly eLearning web log in their case study. The suspicion is these prediction models could be predicted another case of incomplete eLearning web log. According to several pieces of research in designing eLearning, the implementation of eLearning for developing online courses should be considered to meet the high learning effectiveness (Chatteur, Carvalho, & Dong, 2008; Chun Xia et al., 2010; Gang, 2010).

Typically, to approach the high learner outcomes should be constructed by complete web log that could build the learning outcome prediction model for approaching the high learner achievement. Unfortunately, most of the web log is incomplete on eLearning activities because of eLearning implement limitation as for the study of Basha, Umar, and Abbas (2011). They summarized many related kinds of research that work on a restriction of eLearning. The causes of incomplete eLearning are mainly from poor attitudes of eLearning implementation. Moreover, there are some more related activities discovered, for instance negative view of public organizations on digital learning, learner's prior beliefs on conventional learning, barriers in obtaining best face-to-face communication resources, fear of technologies, inadequate equipment, network connections, teacher training, online Learning effective management, IT support, motivational constructs for using virtual education, time and skills needed in adopting new technologies and designed facilities.

In eLearning applications, there is an interest in predicting the outcome of learning and tracking learning processes to validate their learning outcomes. In the past, instructors applied machine learning technique in processing and forecasting. Later, a complex prediction model has been proposed for predicting student outcomes, allowing learners and instructors to access more exciting information.

This complicated prediction technique can group even more similar student sets (Ravichandran & Kulanthaivel, 2015). Data mining has been introduced in the field of education, especially the study of student behavior in the learning environment through the Internet. Automation helps to make complex, inaccessible, and in-depth information more accessible to manage. In many industries, data mining techniques are used to increase efficiency in decision-making and when it is used in education, it can provide better educational services and can improve the learning outcomes and performance of students (Bansal, Mishra, & Singh, 2017). Hence, this research intends to fulfill the learning outcome prediction model that can be widely used for incomplete eLearning systems.

1.2 Statement of Problems

The learning outcome prediction model is based on the learner's ability to learn and think. Personal learning is a significant learning method. An essential part of personal learning is to assess knowledge and learning, along with the presentation of individual learning resources. Predictable learning outcomes and interventions in different learning processes result in different learning outcomes. The essential elements that influence learning are learning styles, learning prediction, and learning interventions (Baolin et al., 2017). eLearning usage model for open sources such as Moodle that can make prediction for learning outcome or learning performance still deficient and cannot be applied in many institutions (Chien Ming, Chao Yi, Te Yi, Bin Shyan, & Tsong Wuu, 2007; Chun Xia et al., 2010; Ribeiro & Cardoso, 2008).

According to eLearning development, several studies show eLearning usage models (Chih Ping & Yi Chun, 2008; Lingyan, Jian, Lulu, & Pengkun, 2010; Milani, Jasso, & Suriani, 2008a). These models were created for predicting learner usage that can affect learners' performance. Nevertheless, it is also crucial to consider on applying these developed models in other higher institutions because most models in the recent year could not be used broadly (Chien Ming et al., 2007; Chun Xia et al., 2010; Ribeiro & Cardoso, 2008).

The use of large data storage and analysis methods is becoming more popular and evolving. In the educational sector, it still needs more development compared to other industries. Learning from eLearning platforms with real-time data mining is a model for predicting accuracy and observing trends in learning outcomes. Traditional data mining and statistics technologies can help you find knowledge or rules that you do not know in advance. In the association of rule analysis, there are three major types of data mining; cluster analysis, classification analysis and predictive analysis. Data analysis with these techniques along with statistical analysis helps explain the relationships of activities in a dataset for more understanding (Peng, Tuan, & Liu, 2017).

According to the literature reviews of eLearning usage models, there are some studies that have approached the models that could be able to predict the eLearning usage which can be a tools for both lecturer and student to know the learning situation and learning outputs for a preferable achievement (Chien Ming et al., 2007; Chun Xia et al., 2010; Ribeiro & Cardoso, 2008). The studying of eLearning model generation with maximum number of possible algorithms still need to explore the automatic process to find better model (Sanchez-Santillan, Paule-Ruiz, Cerezo, & Nuñez, 2016). The prediction model was developed to identify the students who may fail in the exams necessary more evidences to support current discoveries and increase the prediction accuracy (Pan, Xue, Gao, Wang, & Chen, 2018). Nevertheless, most learning prediction models still support eLearning course usage because the primary focus of these studies based on complete eLearning systems. Thereby, most of the previous studies control the most causes of activities in eLearning system that affected the learning prediction result.

On the other hand, if the study focus on incomplete eLearning using, the problem is it could not control all causes to process the learning prediction model. At the same time, the study could not specify which essential purposes affect the high learning outcome prediction. Moreover, the construction process of learning prediction model needs to know the best relationship between causal activities and more learning outcome predictions that could obtain suitable learning prediction models for incomplete eLearning systems. Currently, there are mostly incomplete eLearning have been used in the real world (Zaki, Deris, & Chin, 2002). Today's eLearning focuses on remote services that create a virtual environment for learners and instructors, but the more critical and less developed mechanism is to focus on the individual differences because the ability of the learner is at a different level.

Most of them have difficulty understanding and building on the ideas they learn. However, students are various in their aptitude for each course, background knowledge, learning habits, learning styles, learning motivation, as well as family backgrounds. Those are influence to their learning behavior and learning pattern. As a result, the online learning process has become increasingly important, focusing on enhancing individual learners' knowledge by customization the learning environment to the unique needs of the learner (Gunathilaka et al., 2017).

Therefore, this study solution should approach its cause activities to understand eLearning activities that affected the high learning outcome prediction based on incomplete eLearning system as shown in eLearning activities and high learning outcome prediction relationship (Figure 1.1).

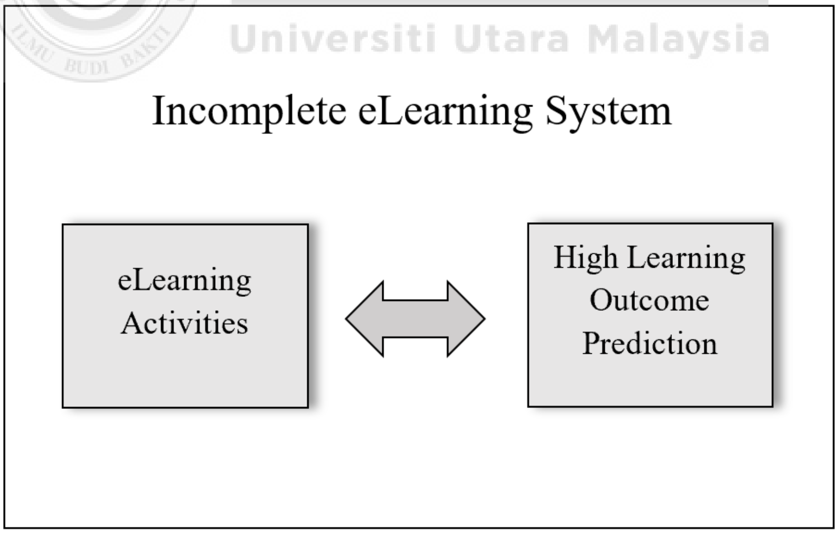

Figure 1.1. Research conceptual framework

The meaning of the word "incomplete eLearning system" in this study is the eLearning system that is chosen to use different tools. For example, homework, forum, quiz, learning resources etc. The selection of tools depends on the needs of the instructors to design those subjects as well as the readiness to use the tools of the learners. In another sense, "incomplete eLearning system" is that the course designer chooses to use only some of the tools provided by the system. Therefore, most courses that are used on the eLearning system will not be able to use all the designed tools. The use of some tools from all here is a definition of the word "incomplete activities".

1.3 Research Questions

Based on the problem discussed, there are few research questions for this work:

1. What are the eLearning activities that affect learning outcome?
2. How can an eLearning outcome prediction model be constructed based on the analyzed eLearning activities?
3. Can an eLearning usage model be synthesized based on incomplete activities?
4. Is the model produces acceptable agreement level on advantage, accuracy, and effectiveness?

1.4 Research Objectives

The main objective of the current empirical study is to explore the activities that affect high accuracy of learning outcome prediction model. For achieving this, the following sub-objectives are as follows:

1. To analyze the eLearning activities that affect learning outcome.
2. To construct a learning outcome prediction model for eLearning usage.
3. To synthesize a dynamic eLearning prediction model based on incomplete activities of eLearning systems.
4. To evaluate the dynamic eLearning prediction model based on incomplete activities of an eLearning system on advantage, accuracy, and effectiveness.

1.5 Scope of Research

The scope of the study are determined to achieve the objectives of the research;

1. The eLearning systems for this study selected from institution which use Moodle open source eLearning systems
2. The data is collected from eLearning web log recorded by an eLearning system from 2012-05-25 to 2015-04-06. There are 257,046 log records.
3. The study is conducted using six semesters' data. There are 53 courses provide by 45 lecturers and 453 students.

1.6 Research Contributions

Based on the objectives of the present research, the researcher aims at making contributions regarding an eLearning usage model for higher education that uses eLearning incomplete activities. The research contributions can be concluded as follows:

1. Analysis of essential activities contributing to high learning outcome prediction.
2. A learning outcome prediction model as a sub model.
3. This study presents the relationship of learning activities that affect to high accurately learning prediction model, and that could be a basis for further research concerning eLearning prediction model.

1.7 Research Significance

In terms of theoretical significance, finding models to predict learning outcomes is a part that helps to understand the hidden pattern of relationships under the vast amount of eLearning usage. Access to a more reliable and more widely used model is a matter of interest in the field of eLearning development. This research will increase the understanding of important activities or processing guidelines that allow access to such effective models.

In terms of practical significance, the model for predicting learning results on the eLearning system that allows instructors to improve the teaching process to be consistent with the situation by applying reliable predictive results as a guide. Students will benefit from showing trends in upcoming academic results in the future.

Such predictions will help learners change their learning habits to better results. Thus creating opportunities for learning to be achieved by the education system.

1.8 Structure of Thesis

This research shows what activities that can affect to the high accuracy of learning outcome prediction model. The organization of this study is as follows

Chapter 1 introduces the research background, gap, problems, questions, and research objectives as well as contributions of the research.

Chapter 2 provides critical reviews of eLearning theories, a brief description of the eLearning usage model, data mining techniques, and data mining in eLearning.

Chapter 3 describes the research methodology which comprising research design, research method, data collection and sampling, model extraction and validation, data analysis, data evaluation, limitation, and delimitation.

Chapter 4 presents the implementation and result of the research that focuses on the data processing result, data characteristics, eLearning activities, data classification, and learning activities.

Chapter 5 highlights the analysis of the research model evaluation for accuracy issues together with an expert review to confirm the model reliability.

Chapter 6 concludes the research and extends discussion on future works as well as the results of the research objectives and suggestions for further study.

CHAPTER TWO
LITERATURE REVIEW

2.1 Introduction

Nowadays, many new technologies are being customized for the developing needs of effective learning including forecasting learner's achievement from learning behaviors via the technological system called eLearning. Apparently, several studies on eLearning development describes eLearning usage models (Chih Ping & Yi Chun, 2008; Lingyan et al., 2010; Milani et al., 2008a), however learner's behavioral patterns collected to predict learners' performance are still limited to a particular user and take more time and effort (Chang, Huang, & Chu, 2009). Therefore, this study makes an effort to construct a suitable and generalized eLearning usage model which originated from eLearning activities relationship in eLearning system environment including system users, system usage procedure, and system tools, called "system usage". Consequently, it is crucial to review and comprehend the eLearning environment.

Over the years, universities around the world have provided educational information and services electronically. eLearning is gaining more attention and acceptance. The key point is that how to enhance performance of the eLearning service including recommender systems to get more learner's attention and acceptance. To facilitate student acceptance and use of eLearning service channels, institutions and eLearning researchers must understand the activities that influence the learner's acceptance and usage of eLearning systems (Alharbi & Sandhu, 2017).

What theory is suitable for the study of eLearning usage? eLearning is comprised of a wide range of disciplines such as education, management, psychology, sociology, communications, library science, information science, social studies of science, social studies of technology and computer science. Therefore, the study of eLearning should assemble these related supporting theories mentioned hereafter. There was a discussion on “Does eLearning need a new theory of learning?” (Haythornthwaite & Andrews, 2011). The result of their discussion is “yes” because of these three reasons. Firstly, they can assume eLearning is somewhat distinct from traditional learning if they are convinced of the assumption “learning is socially situated” which referring that e-classroom society are different those of traditional classroom at schools or universities. Second, digital technology nowadays affects the nature of knowledge itself especially the out-of-relationship leveling among prevailing knowledge, teacher and student. Besides an ordering idea of knowledge, eLearning technologies introduce flatter gradually increasing equality, greater effective dialogical relationship between learner and knowledge. Finally, it is easier for transduction with a multimodal computer interface, whereas the word “transduction” is likely a transformation which is a key feature of learning theory. However, there is still no new eLearning theory. While, researchers in the area are working on enlarging the knowledge base to be the eLearning theory. For this reason, the eLearning study is still based on learning theory.

However, this study focuses on an information technology domain. The new information technology dimension can be conducted with the higher analysis with large attractive data. At the same time, this study data will gain from the web log.

The web log data or usage data of the web is the data set collected after eLearning users did their activities. Thereby, this kind of data set derives from real user behavior. The primary target of this research will be the study of eLearning activities relationship to understand their causes and the results could be applied to new suitable eLearning usage models for students and teachers to get the higher learning achievements. Hence, this study will process through machine learning principal such as data mining techniques and other statistic functions mainly.

2.2 eLearning

Learning can be defined as the pursuit of knowledge or skills through study, experience, or instruction. Learning with eLearning technology is becoming a daily term. eLearning, in other words web-based learning, can be defined as a new changing in the technology of educational technology, which presently has got high intention. Knowledge through eLearning is considerable for educational institutions and corporate training (Alharbi & Sandhu, 2017).

eLearning is a pervasive technology that tries to implement the recommendation to get better impact analysis and to preserve the consequences. Students will use this opportunity to increase their use of eLearning in the same lecture class, as learning is more accessible, faster, more cost-effective, and more reliable. The application of big data with eLearning has made a significant impact on the education system. In the current trend, social media plays a vital role in the eLearning system. Practical use of information is based on the way in which learners and learners use this information. This article has the impact of Big Data on eLearning (Sheshasaayee & Malathi, 2017).

With the emergence of eLearning, governments provide opportunities to learn online, whether formal or informal. However, most Indonesian eLearning systems have been used in the formal educational environment today. Therefore, this study offers an eLearning model to support informal education in Indonesia. It is called as eLearning for the Equivalency Education Program (E-LEEP) model. E-LEEP consists of three components: User, Education Program, and Monitoring. Users are students and tutors. The education program consists of Package A, Package B and Package C for elementary, junior high school and senior high school. The monitoring is used by institutions and stakeholders. Each element supports the needs of students in the eLearning curriculum to achieve their learning goals (Yel & Sfenrianto, 2017).

Presently, the significant roles of eLearning are solving the restraint of space and time problem in conventional classroom to increase the individualized environment in learning and provide a customized knowledge service which can be relied on theories such as modern pedagogy and psychology (Chun Yong & Qi, 2007). eLearning system is well-known for its suitability, adjustable feature and it is a low-cost investment which is now being a new selection to obtain a new feature of education (Xu & Jun, 2009).

In the current technology-driven era, ICT developments have changed dramatically. eLearning has improved the way of teaching, class interaction and the way to assist student in order to provide quality education and flexibility (The & Usagawa, 2017). The key factors that can indicate eLearning readiness compose of several elements for example lecturers, students, technology and environment that must be prompt draw up reasonable and successful strategies. Also, it is likely that the more students spend time with internet-based course the more they get their own learning process and obtain the most significant learning benefits. More than that learning has been developed from a teacher-centered approach to a student-centered approach. Therefore, the crucial component becomes a student who utilize an eLearning course in which their attitude is also very considerable for making eLearning system successful (The & Usagawa, 2017).

One objective of this study is to find the advantages of eLearning. As the work that has been found the special features of present-time education have relied on the use of ICT (Information and Communication Technology) which are gradually raising its importance in term of the teaching of courses with a different emphasis. eLearning has been recently considerable as an powerful technological innovations which can be economical support much more than traditional teaching.(Homiakova et al., 2017). Nowadays, it is very interesting for all existing eLearning initiative to enhance economic performance as a recently rapid growth of eLearning investment worldwide (Yi, Zuo, & Wang, 2007). So, a measurement of investment proficiency assessment or of comparing the ability of various investment is called return of Investment (ROI).

This is a model showing the return of investment (ROI) analyzing and evaluating in eLearning. The highlight of the model puts among the key points that must be mentioned to use the proper investment strategy. Besides, it is further pinpoint the return on the investment as shown in Figure 2.1.

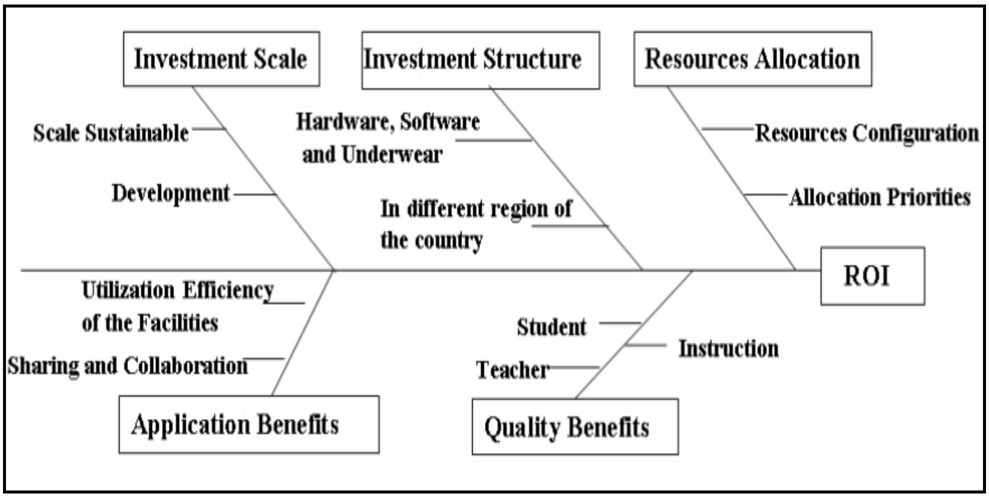

Figure 2.1. The analysis of ROI in eLearning by fishbone diagram

Five important factors that affect ROI evaluation in eLearning can be clarified by this valuable framework in the fishbone diagram representing the analysis of the ROI (Figure 2.1). Besides, sub-factors of each main factor are also described. For more explanation, the five main bones of the spine of a fish represent five factors which branch off into five small bones showing sub-factors. This model aims to analyze and evaluate the return on investment in eLearning by two prominent aspects; the costs and the benefits.

According to Gowda and Suma (2017), seven advantages eLearning are identified as: 1) Web-based learning offers a new mode of interaction between students and students and teacher to student. This mode of education is often called eLearning 2.0 2) eLearning has no geographical borders and is able to spread knowledge to rural areas. 3) eLearning assists colleagues from different work sites to collaborate and share knowledge. 4) It can help to reduce a cost of education. 5) The purposes of eLearning are enhancing web-based learning’s flexibility, scalability and rapid deployability. 6) The innovation in eLearning can create learner's networks which lead to social media-based learning. 7) It helps promoting constructivism of education. (Gowda & Suma, 2017).

The development in eLearning offers an additional method to analyze learning sequencing that analyses in real-time and reports back to learners for choosing learning contents more conveniently (Yen Hung, Juei Nan, Yu Lin, & Yueh Min, 2005). Some developers aim at developing a web-based system for eLearning to teach, to support computer-aided manufacturing, to enhance the quality and quantity of educational technology and to create machine learning for obtaining the web log from eLearning system (Min, 2005).

The eLearning system named "Integrated Distance Education Application (IDEA)" shown in Figure 2.2 was emerged in 2013.

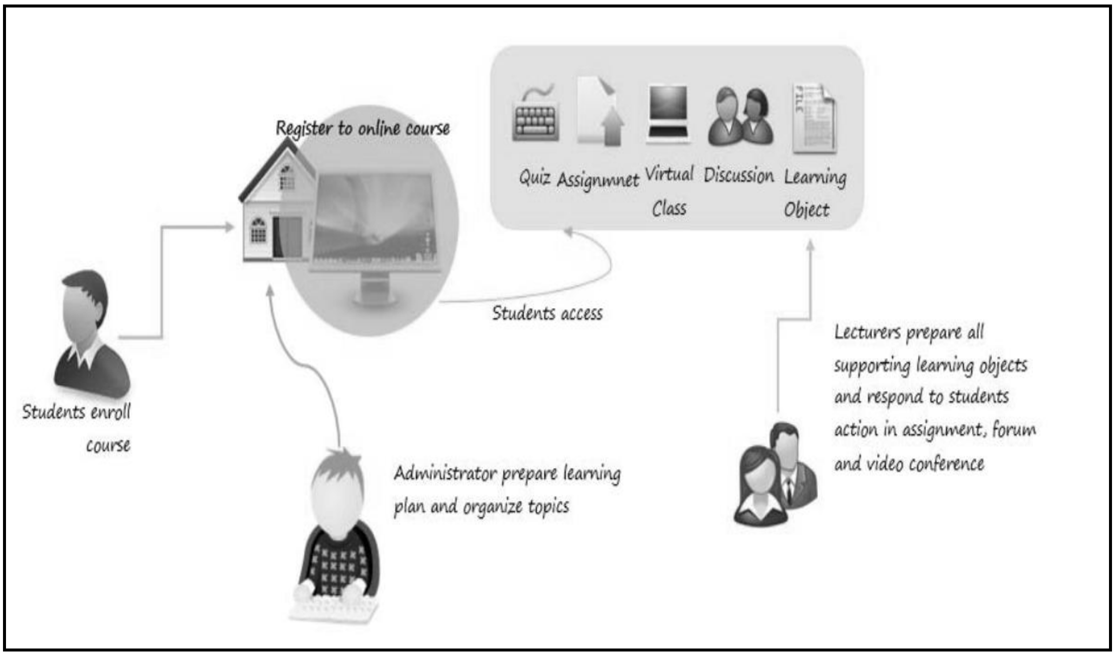

Figure 2.2. General description of existing eLearning system

Figure 2.2. shows the mechanism of the Integrated Distance Education Application (IDEA) which comprises:

- Administrator (or lecturer coordinator) arranges lesson plans and organize module;
- Lecturer synchronizes planning for learning as well as provides any learning tools that is applicable in the form of discussion, course material, assignments, and quizzes;
- Students access the application according to their needs and the applicable rules;
- Lecturers will respond to the actions of the students both in making the assignment and respond to communications from students in the form of

forums and video conference, and students have a right to know the value and the feedback from the lecturer.

This eLearning system has completely combined with the academic. In IDEA, a team teaching with parallel class has adjusted the way they interact with learners. It also allows individual learning to meet their need as well as provides indefinite interaction to other students and to lecturers. Besides, learners can be flexibly grouped by lecturers up to their potential and typical characteristic.

Likewise, another research aims to set up eLearning material for the students and also offer strategies for teaching through the eLearning system. Not only teaching material and strategies, but this study also desires to create the machine learning laboratory to introduce the training data to eLearning system and to combine multimedia to support the effectiveness of machining learning through eLearning system. The method needed for evaluating and analyzing the performance of web application relies on the multi-level web log created by the implementation of web application levels (Xiaokai, Qiuhong, Yongpo, Ji, & Chao, 2010).

eLearning is a high value industry which is around $ 56.2 billion, and this will double in the next few years. Currently, more than 40% of companies worldwide utilize some sort of educational technology to train employee. Consequently, more attention was gain to the Learning Management System (LMS) to systematically implement, make improvements and manage eLearning, whether the institution provides primary and secondary education, higher education, continuing education, or professional industry training programs. Continuous development of LMS improves access and tracking of learning activities. It also supports the organizational growth and development. Connected to present social cloud networking trends the next generation of LMSs will be opened, mobile, personal, social, flexible and will have an additional tool for learning analytics. This new generation of LMSs as a new study methodology that must be able to meet the needs of the changing environment of business and education (Paľová, 2016). Besides, either commercial or open source eLearning system was applicable in the majority of an educational institution in which apply Learning Management System (LMS) which mentioned hereafter.

2.3 eLearning platforms

In modern education are using Learning Management System widely. In traditional education, the instructor does not have a thorough understanding of how his students are using educational resources. However, the situation will vary for the online learning environment. Web logs contain a variety of information about the visited pages. These web logs become a rich source for data analysis. Understanding the usage pattern from web logs is widely used to improve eLearning system (Dragoş, Săcărea, & Şotropa, 2017).

Educational data is abundant due to the nature of 21st-century education, where many studies have taken place online through the advent of digital learning. The use of Learning Management Systems (LMS) has increased their significance to provide education without boundaries, and these systems hold a massive amount of data. Educational data mining is a process of accessing a wide range of educational data, with the objective of developing the education process through the knowledge gained from the mining process. A paper is intended to be a link to a master's degree course that has been excluded from Moodle LMS. This point may lead to a decision that can be made to improve the learning process. The results show that most students have access to the resources available during the last minute before the exam and are likely to procrastinate online submissions (Nkomo & Nat, 2016).

One of the most popular LMS nowadays is LMS Moodle which is able to support highly in daily teaching. LMS focuses on supporting ICT-based education. This structure is influenced by the teaching hypothesis and the content plan of each subject. Presently, these electronic education products focus more on the personality of the learner. The most appropriate method is to analyze its properties to determine the advantages and disadvantages of LMS using. For this purpose, they decided to use strategic SWOT analysis to present this finding as strengths and weaknesses external factors affecting the system concerning possibilities and threats (Mudrák, 2017).

Strengths:	**Weaknesses:**
• free of charge • used at our university • intelligibility • multiplatform • the multilingual system including the support of the Slovak language • support of adaptive elements • simple administration • evaluation of students • support of group activities • open-source • plug-ins • an online community • regular upgrades • the most used LMS in Slovakia	• high upload limit • surveying study materials • limited functionality • necessary hosting • need to connect to the Internet
Opportunities:	**Threats:**
• students are comfortable using it • introduction of the subject of eLearning technology • creation of work methodology with an adaptive LMS	• the inexperience of students working with an eLearning course • students not willing to cooperate in the research • insufficient readiness of teachers to work with an adaptive system

Figure 2.3. Learning Management System (LMS) SWOT Analysis

Learning to achieve effectiveness and advantage of eLearning is part of this study objectives. From Figure 2.3, comparing the advantages and disadvantages of eLearning with the SWOT Analysis method, which will see the importance of the eLearning system that results in the development of the learning system.

The Learning Management System (LMS) is a software program for document management, tracking, reporting, and delivery of eLearning courses or programs. Using LMS in the educational institution can help teacher's attention to the needs of students. The online access to all data needed for learning has an immense benefit not only for students but also teachers (Paľová, 2016).

In education system, more attention has been paid to learning management systems along with new technology. By using Moodle to record user experience in eLearning systems in the learning management system, Moodle provides information exchange and communication among participants in eLearning courses and analyzes the user experience to meet the teaching needs of both students and teachers in all areas. Using Moodle will make learning more accessible and more interesting. Moodle Learning Management Systems will help academics build productive online learning communities using data mining techniques. Moodle log files help the instructors to preprocess the data, predict learning strategies and summarize the website structure according to learner's interest by applying data mining techniques (Sheshasaayee & Bee, 2017).

LMS provides a variety of features for managing structured learning. These features are integrated with academic registration, plan learning, learning content access, online discussions, online testing and monitoring the progress of the learning process (Laksitowening et al., 2016).

A learning management system (LMS) is application software used for administration, documentation, tracking, and reporting of training programs, classroom and online activities, eLearning programs, and training. Moreover, LMS is an efficient system, which has various abilities ranging from systems for operating, training and recording educational activities to software for setting up courses via the Internet with various features for online cooperation. A training section in a corporation uses LMSs to automatically keeping a record and registering employee profiles. There are several dimensions to Learning Management Systems (LMS) (whatislms.com, 2012). Student self-service such as self-registration, training workflow such as user notification, manager approval, and waiting-list management, the preparation of eLearning such as Computer-Based Training and reading & understanding, online assessment, management of continuous professional education (CPE), learning collaboration such as application sharing and discussion threads, as well as management of training resource such as instructors, facilities and equipment.

The use of LMS has grown over the years, so in countries such as Germany, most universities use such systems in both direct and distance education. They are mainly used in schools, colleges, universities and even in some organizations, such as support systems in many areas, especially for sharing knowledge. As LMS has been considered as a source of education data, processing methods for this information have been created. The general analysis of online learning is based on assumptions as an instructor or administrator who assumes the role of a miner makes questions and uses the analyzed data to answer that question (Nkomo & Nat, 2016).

LMS can log all student hits in the log file, which includes the students' view of course information, the students' activities and also the interactions while using the system. One of the popular data mining technique is clustering which can be used to analyze students' behavior in LMS. In the study, clustering helps to divide the data into groups of students depending on some categories such as learning materials preferences, learning behavior, and learning environment improvement (Alias, Ahmad, & Hasan, 2017).

There is one article shows the case study of the analysis of the log of Learning Management System (LMS) in the flipped room. They examine the relationship between the preparation of students before class and their test scores through LMS log. Case studies provide students with papers and quizzes about the content of the next class through LMS. Then, a simple test is conducted in the first 10 minutes of the class to evaluate the students' understanding of content. They pay much attention on the amount of times a student reads a document and takes the quiz. This article analyzes the relationship between student preparation for the next class and their test scores using Moodle logs. They can find some patterns that show the characteristics of some students in one score category (Yamada & Hirakawa, 2015).

Paľová (2016) summarized the common features of a Learning Management System (LMS) as:

- Content authoring/resource management gives the ability to users to design their content and deliver courses within an LMS. Authoring and sharing tools are often part of an LMS.
- The ability to control the user's activities. The LMS provides a set of tools and rules for setting permissions of access, setting up activities, determining to share, and initiation of a study group.
- Proficiency testing and reporting help administer tests to measure employee/student knowledge or skill. Analytics and reporting functionality can identify learning gaps.
- Learning management tools helps to organize and simplify training or learning administration, including distribution of content, management of user information, schedule and oversee course enrolment. Learners can set learning objectives, monitor, and evaluate their learning process and achieved knowledge.
- Course and the personal library provides a pre-made library of training courses for general purposes, such as those on sexual harassment policies or management techniques. At the same time, LMS provides a place for the user to create and publish various kind of content via blog, wiki, forums.

- Certification and compliance management including setting up, tracking and managing certification programs for industries that require employee certification.
- Virtual classroom via functionalities like video conferencing, live course/lesson leading, remote classes through the platform.
- Extended enterprise means that LMS allows organizations to train or teach external users. For that purposes, the E-commerce functionality may be included.

Furthermore, Learning Management System (LMS) is a learning system software that shows theoretical content in an organized and controlled way. It mainly consists of administration, content packing, synchronous and asynchronous communication tools, knowledge evaluation and tracking users (Sancristobal et al., 2010). Besides, LMS offers a platform to create interactions among students and tutors, and among the peers. In an eLearning environment, it can show most of the traditional pedagogic activities (Hsien Tang, Chih Hua, Chia Feng, & Shyan Ming, 2009). Accordingly, LMS has been developed to replace real classrooms. Nowadays, higher institutions look for the best LMS, either open source that cost nothing or closed source for commerce. However, most LMS are developed to meet the standard pattern that can use with other eLearning systems such as Shareable Content Object Reference Model (scorm.com, 2012) that is a gathering of standards and detailed description for eLearning and it is the most widely used standard pattern.

There are five modules in LMS; course management module, authoring module, collaborative learning module, assessment module and administration module which are distinguished by the eLearning environment oriented design for individualized adaptability study (Ninomiya, Nakayama, Shimizu, Anma, & Okamoto, 2007). Hence, these modules will be the sources of data collection in this study.

Moreover, another advantage of open source eLearning is modifiable resource systems that researchers can extract interesting data from databases including web logs. Furthermore, the different analytical purposes, from basic approaches such as statistical information about accessing the platform can use this data for complicated information about user action sequence with all related information (Babic, Wagner, Jadlovska, & Lesko, 2010).

2.4 Open Source eLearning

Recently, many open source eLearning platforms were evaluated by Graf and List (2005) which consisted of ATutor, Dokeos, dotLRN, ILIAS, LON-CAPA, Moodle, OpenUSS, Sakai and Spaghetti learning were focused on adaption issues. The result of the comparison in Table 2.1 shows that the platform Moodle exceeds all other platforms and also gained the highest rating in this category. They classified all activities into eight subcategories that can be shown in Table 2.1.

Table 2.1

The result of Adaptation Category (Graf & List, 2005)

	Adaptability	**Personalization**	**Extensibility**	**Adaptively**	**Ranking**
Maximum values	*	#	*	*	
ATutor	|	#	#	|	3
Dokeos	|	0	*	+	2
dotLRN	+	+	*	0	2
ILIAS	+	#	*	0	2
LON-CAPA	+	#	#	|	2
Moodle	#	+	*	|	1
OpenUSS	#	#	#	0	2
Sakai	0	0	*	0	3
Spaghettilearning	+	#	+	0	3

The symbols meaning of Table 2.1: E = essential, * = extremely valuable, # = very valuable, + = valuable, | = marginally valuable and 0 = not valuable

Table 2.1 shows the comparison of many eLearning features that are widely used around the world, including ATutor, Dokeos, dotLRN, ILIAS, LON-CAPA, Moodle, OpenUSS, Sakai and Spaghettilearning. Issues used to compare the different features of each eLearning are adaptability, personalization, extensibility and adaptively. Comparison results with all four aspects shows that Moodle is an eLearning system that has a best ranking by average. In this study, Moodle eLearning is used as a case study.

Table 2.2

Activities of Open Source eLearning (Graf & List, 2005).

Subcategories	Activities
Communication tools	Forum, Chat, Mail/Messages, Announcements, Conferences, Collaboration, Synchronous & asynchronous tools
Learning objects	Tests, Learning material, Exercises, Other creatable LOs, Importable LOs
Management of user data	Tracking, Statistics, Identification of online users, Personal user profile
Usability	User-friendliness, Support, Documentation, Assistance
Adaptation	Adaptability, Personalization, Extensibility, Adaptively
Technical aspects	Standards, System requirements, Security, Scalability
Administration	User management, Authorization management, Installation of the platform
Course management	Administration of courses, Assessment of tests, Organization of course objects

Table 2.2 (Continue)

Table 2.2 mainly presents activities that appear mostly in open source eLearning systems. According to this study, the focus is put on eLearning usage, so that user web log will be collected from Moodle open source eLearning to construct eLearning user behavior model because it is in the highest ranking open source in term of an adaption issue.

2.5 Moodle

The development of Internet functions has influenced many activities, conversations, meetings, shopping, and learning. eLearning is part of a development model of education that uses the Internet. The current development model of online education still needs support and innovation. Moodle is a learning tool used in education, administration with a robust, secure, and integrated system to create a personalized learning environment. Free web-based open source PHP applications for the production of modular internet-based courses that support modern social constructionist pedagogy. M in Moodle stands for modular. Building on new functionality in Moodle by writing this kind of plugins is the easiest and the most common way. Plugins available in Moodle are various such as activity modules, reports admin, admin tools, reports, plugins, themes authentication, and more (Susanto, Irdoni, & Rasyid, 2017).

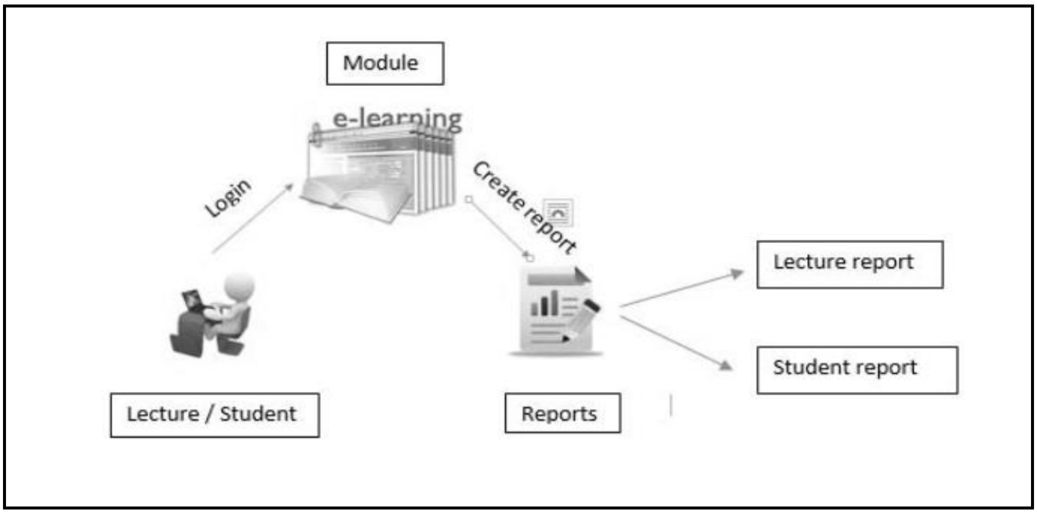

Figure 2.4. How plugin works (Susanto et al., 2017)

Figure 2.4 shows an idea about the functionality of the plugin. When the user logs into an eLearning application, new information is added to the database, which contains the values to the time that the user logs in. In addition to the user's login data, this plugin records user activity in the course. Besides, based on available data, the plugin provides a two-part report, reports that report data on user role teacher, roles, and students' user reports. For this study, using the same approach as Figure 2.4 in presenting modules that can be installed into the eLearning system. When the module is working completely, the report will be sent to the instructor and learner as well.

Because Moodle design is made by social constructionist pedagogy, which focuses on a learner-oriented philosophy, it called modular object-oriented dynamic learning environment (Moodle) that the best LMS. More than 200 countries have widely adopted, more than 40,000 sites registered, and more than 2,400,000 courses (Hsien Tang et al., 2009). Using Moodle will make learning more accessible and more interesting. Moodle Learning Management Systems will help academics build active online learning communities using data mining techniques. Moodle log files help the instructors to preprocess the data, predict learning strategies and summarize the website structure according to learner's interest by applying mining techniques (Sheshasaayee & Bee, 2017).

Moodle is not only a LMS or a Virtual Learning Environment)VLE((Moodle.org, 2011) but it is also a Course Management System (CMS). This is a free web application for creating useful online learning sites and also included Sharable

Content Object Reference Model standard ready for educators (SCORM) (Ruiz Reyes et al., 2009; scorm.com, 2012).

Likewise, Moodle is a software package for producing Internet-based courses and websites. This development project is designed to support a social constructionist framework of education globally. Moreover, Moodle is provided freely as Open Source software (under the GNU Public License). Basically, this means Moodle is copyrighted, but that users have additional freedoms. It is quite user-friendly in which is copiable and adjustable by users. Moodle provided that user agrees to give the source to others; not change or remove the original license and copyrights, and apply this same license to any derivative work. Finally, any kind of machine that can run by PHP or SQL for example MySQL can run Moodle. Besides, Windows and Mac operating systems and variations of Linux (for example Red Hat or Debian GNU) can also use to run Moodle as well. There are many knowledgeable Moodle Partners to assist users and event host Moodle site.

Moodle's logs data, combined with detailed data from other system components, will allow teachers to determine the intensity and type of students' activities to analyze their test and assignment results or to provide a complete report of individual students or of students groups for a specific activity. They have studied case studies of pre-processing operations, which make raw data collected by the Moodle system in the “Stefan cel Mare” University of Suceava, in a dataset that permits the use of mining algorithms. The obtained format is intended to predict students' expectations, outcomes based on the type of questions, and to search for correlations between the results of the study and the courses that are linked together by topics covering most operations being performed. In SQL, the advantage of the facilities offered by the database management system (Danubianu, 2015).

Generally, the usage of this open source software provides the evolution of learning tools and increases the quality of education. The study of analyses and comparisons has been conducted about open source learning management systems and Moodle with many outstanding features among other LMS that aims to enhance the quality of education and combine the tools that a system should have (Aydin & Tirkes, 2010).

"Moodle" is the one outstanding eLearning system (Aydin & Tirkes, 2010). It also includes the famous eLearning standard feature SCORM. Shareable Content Object Reference Model (SCORM) is an XML-based framework used to define and access information about learning objects which is conveniently sharable in the thick of various types of Learning Management Systems (LMSs). For promoting a standardization in eLearning, a United States Department of Defense (DoD) initiated an application of SCORM. The DoD had been frustrated by problems they encountered when trying to share eLearning courses to various LMSs that were used in that organization. Therefore, they formed the Advanced Distributed Learning (ADL) specification group was created in 1997 as a tool to promote a way to set learning content portable across various systems. ADL created the first version of SCORM, which initially was applicable for SCORM. To assist the flow of course content and related information (e.g. student records) from one platform to another, to make course content into modular objects that can be reused in other courses, and to enable any LMS to search others for available course content, ADL is designed for those mentioned purposes. (scorm.com, 2012).

The SCORM identifications, which are distributed through the Advanced Distributed Learning (ADL) Initiative Network, define an XML-based means of representing course structures, an application programming interface (API), a content-to-LMS data model, a content launch specification, and a specification for metadata records for all components of a system. The ADL specification group's next challenge is to motivate vendors to comply with SCORM specifications.

SCORM is a set of courseware standards of eLearning from the United States, which have a significant impact at the international level. The description of the SCORM Data Model is abstract; it is difficult to understand and master for ordinary users and developers. Based on in-depth research on SCORM RTE (Run-Time Environment) model, JavaScript and the logical performance of SCORM Data Model set up a detecting procedure of Data Model which can observe on the Moodle platform. At that time, the interactive discipline between Learning Management System and SCORM Data Model is displayed directly, and the application character of data parameters are described in detail. It is such a great assistance for designing a Learning Management System or developing network courseware by using SCORM criterion (Wang, Lu, Zhao, & Cha, 2010). A SCORM-compliant learning management system (LMS) has been developed that enhances learning by effectively and efficiently managing the learning itself. First, a SCORM-LST was developed by adding to SCORM (shareable content object reference model) a framework that describes facilitation corresponding to the learner's state of learning and describes the learning state transitions (Morimoto, Ueno, Yokoyama, & Miyadera, 2007).

In the process of using developing Web-based training (WBT), SCORM is the most common eLearning standard that is able to integrate eLearning objects from different sources in a classic environment. However, computer-based training (CBT) does not aim to deliver offline on CD or DVD. It can be gained from accessing to eLearning market by SCORM as well as eLearning course creators could probably be advantageous from applying e technology for online and offline publishing (Nordmann & Neumann, 2008).

For this SCORM reviewing, many standard table attributes represent as usage activities. This study uses this aspect as a basic structure of the eLearning system includes the learner's activities that were used in this study processes.

According to the study of Ribeiro and Cardoso (2008), the data gained from the web log in Moodle is not only necessary as a navigational framework but also provides related input for a cautious model construction which is needed for tracking students' behavior. The result illustrates the model, which can successfully forecast students' final performance while bringing useful feedback during course flowing.

This study uses an online learning program using Moodle LMS. However, LMS requires a plug-in that monitors student activity, which is when using the application. Moodle plugins must be developed to monitor student status. This plug-in has a function to record the user's login time as a marker of the presence in eLearning. The data is accessed using the PHP code added to every page in Moodle, in addition to observing the value of the logged-in user. This plugin also records some activities performed by the user. Besides, the plugin will process the information and display it as a report with the appropriate format, as defined by PENS. This plugin has been validated for its ability to provide information about user activity when using the eLearning application (Susanto et al., 2017).

Reviewing this topic, Moodle is an open-source that became a famous system because of its performance and popularity (Ribeiro & Cardoso, 2008). So, as shown in the methodology of this study, the Moodle system web log will be used in this study.

2.6 eLearning Usage

According to the definition of eLearning behavior defined by Kebin, Feimin, Ming, Feng, and Xiaoshuang (2008), "eLearning is the long-distance independent learning behavior that takes place in the learning environment which was constructed by information technologies. A learning portfolio is an eLearning usage data that provides the students with a specific method for evaluating their own learning situations". The records of the students' activities during the learning process, such as their interaction with others, assignments, test papers, personal work collections, their discussion content, and learning records are all included (Chien Ming et al., 2007). The structure of the database from the system illustrated below; both tables derive from open source eLearning system in Moodle. Table 2.3 and Table 2.4 present the group of tables that contain data for the data mining process.

Table 2.3

Some Important Moodle Tables for Data Mining (Romero et al., 2008).

Name	Description
mdl_user	Information about all the users.
mdl_user_students	Information about all students.
mdl_log	Logs every user's action.
mdl_assignment	Information about each assignment.
mdl_assignment_submissi	Information about assignments submitted.
ons	Information about all chat rooms.
mdl_chat	Keeps track of which users are in which chat
mdl_chat_users	rooms.
mdl_choice	Information about all the choices.
mdl_glossary	Information about all glossaries.
mdl_survey	Information about all surveys.
mdl_wiki	Information about all wikies.
mdl_forum	Information about all forums.
mdl_forum_posts	Stores all posts to the forums.
mdl_forum_discussions	Stores all forums' discussions.
mdl_message	Stores all the current messages.
mdl_message_reads	Stores all the read messages.
mdl_quiz	Information about all quizzes.
mdl_quiz_attemps	Stores various attempts at a quiz.
mdl_quiz_grades	Stores the final quiz grade.

Table 2.4

Attributes Used for Each Student (Romero et al., 2008).

Name	Description
course	Identification number of the course.
n_assignment	Number of assignments handed in.
n_quiz	Number of quizzes taken.
n_quiz_a	Number of quizzes passed.
n_quiz_s	Number of quizzes failed.
n_messages	Number of messages sent to the chat.
n_message_ap	Number of messages sent to the teacher.
n_posts	Number of message sent to the forum.
n_read	Number or forum message read.
total_time_assignment	Total time spent on assignment.
total_time_quiz	Total time used in quizzes.
total_time_forum	Total time used in forum.
mark	Final mark the student obtained in the course.

A collection of standard web log could be beneficial as it helps to clarify the things were happening in eLearning usage. Also, it explains the eLearning usage in different periods of time and different user groups. Studying the eLearning usage is useful.

The process of data mining is analysing eLearning usage web log to introduce the learner's dynamic information. Web data mining is used to process in detail and explain individual learner information, model's establishment and the updating process (Luo & Chen, 2010). Besides, data mining can also describe a mixture model-based approach to generate and visualize individual behavior models for the users and represent the web log as a collection of the ordered sequent action for each user (Manavoglu, Building, Pavlov, & Giles, 2003).

In web-based learning, an assessment of the learner's performance using portfolios is now a prevalent issue in conducting research. The study result reveals that the evaluation of the offered design is similar to those of accumulative assessment results of grade level. Moreover, teachers could try to be more understanding of the activities influencing learning performance in an eLearning environment (Chih Ming, Chin Ming, Shyuan Yi, & Chao Yu, 2006).

Moreover, Munoz-Organero, Munoz-Merino and Kloos (2010) have analyzed the relationship between student motivation and performance and collaborated student's interactions with eLearning system (Munoz-Merino, Kloos, & Munoz-Organero, 2010) that can forecast the shortage of motivation on Learning Management System Moodle.

Also, Yongquan, Zhongying, and Qingtian (2007) conducted the study on eLearning interest of a user to analyze the reading behavior of a user by focusing on learner's actions such as underlining, highlighting, circling, annotation making and bookmarking on e-document during their studying. It reveals the proposed behavior table to record these behaviors.

According to the study on attitude and behavior, Lui (2005) asserts that the adoption of technological innovation for a faculty is a result of having a positive attitude toward technology. Moreover, there is a reciprocal relationship between attitude and behavior in which each one affects the other, that is, a positive attitude can stimulate teachers to apply a particular technology, and positive experience resulting from using that technology can reinforce the already established positive attitude.

An individualized learning strategy in a smart eLearning environment has described personality characteristics of the learner, strategy character of learning and behavior factor, which satisfies the present request of individualized online learning. According to the process of personality analysis, it can acquire the relationship between behavior characteristics and personality characteristics, and the relationship between personality and strategy. After that this intelligent eLearning systems can apply the relationships for solving problems related with individual learning strategy and making better learning, and implementing personalized learning strategy (Feng, Qinghua, Zhiyong, Jin, & Renhou, 2007).

Feng-jung and Bai-jiun (2007) presented the study on the material using activities for finding out the patterns from the vast amount of learning web log of user and for accessing the accuracy of recommendation system for learning material. Moreover, it is shown that the material recommendation system design is under the basis of the learning activities of the previous group of learners which proposed an integrated learning activity-based mechanism to support users with material recommendation automatically.

Learning styles are often unique. In the present situation, many educators, psychologists, researchers are beginning to recognize different learning styles. Before learning the subject, understand and recognize the learner's learning style to enhance the learner's performance. In-depth understanding of the learner is vital in eLearning. It is important to have a clear perceptive of the different learning styles to predict the learning patterns of different learners in an eLearning environment. Learning styles can be predicted according to the learner's characteristics (Deena & Raja, 2017).

2.7 eLearning Usage Model

The one study presented here is dealing with sequential web patterns, which could serve a specific user-specified minimum support. This study also describes the most frequent access sequential relationship of the web page. Besides, the study tried to reveal the ordering of eLearning functions used in eLearning environments by selecting the four most popular functions, which are downloads of courseware, course notice, homework submission and course discussions for making the access priority order. This research also shares some aspect of eLearning behavior studying methods (H. yan Wu, Zhu, & Zhang, 2009).

Moreover, Ting (2008) showed how the data mining used in the discovery-driven exploration of data cubes and presented the contribution to some individual visitors in some community. Thereby, it shows that the construction of Internet education resources can apply advanced technology computer to data mining.

Whereas, Hu (2008) designed and made an application for an evaluation system for eLearning performance. This study considered the majority of the activities involved in an online learning process, the information collected from the evaluation of learning processes and based on the theory of fuzzy mathematics. Hu classified the results of the learners in online learning into five categories; Excellent, Good, Fair, Pass, and Fail. They are considered to be fuzziness of the evaluation results, and besides, it is supportive of student motivation in online learning.

Eom (2010) studied students satisfaction learning and students outcomes for the university's online education via eLearning. He introduced this research for improving the effectiveness of eLearning system and for comparing the efficiency of the learning outcomes between eLearning system and face-to-face learning system. Therefore, he presented the model, which comprises three independent variables; system quality, information quality, and self-efficacy and four dependent variables; system use, user satisfaction, self-regulated learning behavior, and eLearning outcome.

Solak and Cakir ((Solak & Cakir, 2015) focuses on the language learning strategies used by learners and discusses the correlation between academic achievement and these strategies. They find that students should be encouraged to participate in the eLearning program to learn foreign languages. The flexibility of the eLearning program may be a reason for this. In addition, learners take advantage of metacognitive and memory strategies more frequently than other strategies. This will lead to the enrichment of the eLearning program and promote the eLearning program for other disciplines. According to the study of English eLearning in the virtual classroom and the activities that influence ESL, they find that most English learners think that using the internet to learn English is more convenient than using traditional methods. And most English language learners think that using the Internet to learn English is more useful than using traditional methods (Tan, 2015).

Moreover, differences between the characteristic patterns of student learning behavior in online learning are also an attractive area of such work (Tsung Ho, Kun Te, & Yueh Min, 2008). From this study, there is a two-stage cluster analysis established and designed to review the identification and categorization of learning behavior patterns of the students.

For this eLearning usage model discussion, it demonstrates the practical modeling base on eLearning user' behavior model development. Many aspects of eLearning usage studying have been finding the appropriate solution to construct the better eLearning usage model due to their purpose. According to these reviews, this study should put more focus on web log specification, eLearning usage activities classification.

The learning style of the students influences the learning process, so it determines the learning achievement. To accommodate the differences that may occur with students, they need to use a personal learning process. The research presents an eLearning system that is tailored to the personalization based on the Felder-Silverman Learning Style Model. The learning style is identified through a questionnaire and determines the sequence of learning and learning objects recommended for each student. This research offers an ontology model to support personal adaptation in eLearning. Ontology can show the objects and relationships required to implement a learning model and to introduce learning objects. This research has to be validated through experiments to find out what makes people more involved in learning outcomes (Laksitowening et al., 2016).

Data mining technique plays an important role as it helps to extract the information from a vast database called weblog. According to the study of Wen-Hai (2010), the study of using data mining technique to bring out behavior pattern of the client from web log files which focus on analyzing client's behavior pattern recognition system and its application to obtain client information conveniently and automatically.

Furthermore, the study of behavioral patterns model in eLearning environments suggests a behavioral learning model which is based on Colored Petri Nets (CPN) to form a model and generate students' behavioral patterns. This study results show how to generate pattern of behavioral approaches to actual student behavior and also show the generated behavioral pattern serves as adequate test data to test whether the predicted learning content of a smart eLearning system is appropriate; The suggested model can recommend the proper learning content for students who learn via eLearning systems (Chih Ping & Yi Chun, 2008).

The analysis affords the possibility of the transmission of materials and social culture through language in the cultural arena where cultural artifacts in the agency. In the study of cultural studies, algorithms signal changes from the center of personal or social concerns and into the complex relationship between humans and non-human entities that are prevalent in our digital network activities (Jandrić, Knox, Macleod, & Sinclair, 2017)

eLearning courses are very demanding in recent times. The need to study student achievement and predict their performance will also increase. With the growing popularity of educational technology, data mining algorithms that are well suited for predicting student performance have been reviewed. The best procedure depends on the nature of the faculty's predictions. As the amount of student data increases, the need to handle data complexity and processing becomes a challenge for students at risk. It covers the Decision Tree Approach for Predictive Analytics of student’s performance and its Big data implication (Vyas & Gulwani, 2017).

It is apparent that these studies collected usage from system web log so that it is also crucial for this study to collect data from system web log to get an appropriate usage model.

2.8 eLearning Usage Evaluation

The main starting point for implementing eLearning is to evaluate the readiness of the university and evaluating of resources and constraints. It can be defined as a readiness assessment that the institution will implement and implement eLearning. In consideration to eLearning readiness, it involves many components such as students, lecturers, technology, and environment, which must be ready to formulate a coherent and achievable strategy (The & Usagawa, 2017).

Student performance is an important measure for evaluating the effectiveness of the eLearning platform. But the problem arises that it is not unique for evaluating the quality of eLearning methods; other activities also play an important role. Performance assessments in the context of most eLearning are conducted using online tools to examine the effectiveness of learners according to their knowledge and cannot identify the natural differences that affect the effectiveness of the learner's performance (Al-Alwani, 2014).

After exploring the literature review, they find that students like studying eLearning increasingly. eLearning plays a vital role in the field of modem education. eLearning encourages teachers and students to take personal responsibility for their learning. From several activities to improve the quality of education initiation, the most critical activities is the interaction between technology and human. How can technology transfer knowledge to humans? Many people find that learning experience by eLearning. People who study with eLearning continue to interact in the social community (Wardaya & Pradipto, 2017).

The importance of eLearning is critical to understanding the different learning styles to predict the learning patterns of different learners in an eLearning environment. Learning styles can be predicted according to learners' characteristics. In general, there is a variety of learning styles. The learning style of the learners increases the efficiency of the learners (Deena & Raja, 2017).

According to the study of behavior activities evaluation (Lingyan et al., 2010) it reveals that the attributes in the web log could be classified into three groups as shown in Table 2.5.

Table 2.5

Group of Web Log's Attributes Processing (Lingyan et al., 2010)

Group of Attributes Processing	**Attributes' Activities**
Count	- The number of learning resource (TotalCount) - The number of asking questions (QuesCount). - The number of answering questions (AnsCount). - The number of sending posts (bbsSentCount). - The number of replying post (bbsAnsCount). - The number of tests has been done (TestCount). - The number of assignments (HomeworkCount).
Time	- The average time of learning resource (TotalTime).
Score	- The average score of the tests has been completed (Test Score), which is classified into five classes: 'A' is the score greater than or equal to 90 points, 'B' is the score between 80 and 89 points, 'C' is the score between 70 and 79 points, 'D is the score between 60 and 69 points, 'E' is the score smaller than 60. - The average score of assignment (HomeworkScore).

Table 2.5 presents the learner activities (web log's records) processing types that processes by activities counting, activities timing (period) and activities score. From the starting time of the eLearning system, all of the usages become the activities that can classify into three groups. Hence, these three groups of processing could be used for attribute specifying in data preparation step in this study.

With the emergence of eLearning, governments provide opportunities to learn online, whether formal or informal. However, most Indonesian eLearning systems have been used in the formal educational environment today. Therefore, this study offers an eLearning model to support informal education in Indonesia. It is called as eLearning for the Equivalency Education Program (E-LEEP) model. E-LEEP consists of three components: User, Education Program, and Monitoring. Users are students and tutors.

The education program consists of Package A, Package B and Package C for elementary, junior high school and senior high school. The monitoring is used by institutions and stakeholders. Each element supports the needs of students in the eLearning curriculum to achieve their learning goals (Yel & Sfenrianto, 2017).

Ozkan and Koseler (2009) had studied the evaluation in multi-dimension of e-based learning system for higher education setting by using a Hexagonal eLearning Assessment Model (HELAM) suggested for Learning Management System (LMS). The six proposed factors are Instructor Quality (Factor 1), Information Content Quality (Factor 2), System Quality (Factor 3), Service Quality (Factor 4), Learner's Attitude (Factor 5) and finally Supportive Issues (Factor 6) for instrument survey. The study aimed at proposing a model of eLearning evaluation including a set of measurement of an eLearning system and testing the construction of the conceptual model by using a survey to show how useful an eLearning system is especially implementing for a computer course. Therefore, the survey demonstrates that six-dimensional factors of the model of conceptual eLearning evaluation were significant to utilize in computer literacy course via eLearning management.

The research, namely, "Use Web Usage Mining to Assist Online eLearning Assessment" (Ling, Xin, & YuanChun, 2004) presents a group of models relevant to eLearning activities. Their model consists of the courseware, courseware pages, estimation of courseware page and transaction page for approaching to the behavior of student assessment via data mining.

After the model finding process, output model evaluation step is necessary. Castillo, Gama and Breda (2006) conducted the study on an adjustable predictive model for students which consists of determining what kind of learning resources are more appropriate to particular learning styles. Finally, they have evaluated the eLearning predictive model by measuring the accuracy of the model's predictions on a set of artificial datasets that represents a simulated student. Therefore, the process of eLearning predictive model evaluation could be used for the new dataset for model accuracy. Hence, these study also use the new dataset in making an accurate eLearning usage model.

According to measurements of learning based on eLearning, web log represents the usage that recorded by eLearning system (see in Table 2.3 and Table 2.4). According to the study, it should follow the web log structures that represent the user's learning behavior information to approach the eLearning usage model. In recent years, many algorithms in data mining for solving different problems have been created, such as neural networks, Support Vector Machine (SVM), Bayesian networks, genetic algorithms, fuzzy techniques, and swarm intelligence (Baesens, Mues, Martens, & Vanthienen, 2009) .The optimal detection of patterns in data is the critical focus of these developments, provided a predetermined performance metric, such as an accuracy of classification, R-squared or mean squared error (Baesens et al., 2009).

2.9 Data Mining in eLearning

Online learning has brought the need for a database to store the vast amount of data produced in the education environment and this is how educational data mining was born (Nkomo & Nat, 2016). In data mining, statistical techniques, mathematics, artificial intelligence and automatic learning are the processes used to distinguish and identify useful information from large databases to generate knowledge. Noted that the use of data mining has many formalities that need to be taken into consideration to represent the data; probability, rules, trees and a set of statistical methods. These characteristics help stakeholders save time on tasks such as finding one or a group of individuals with similar groups (Villegas-Ch & Luján-Mora, 2017).

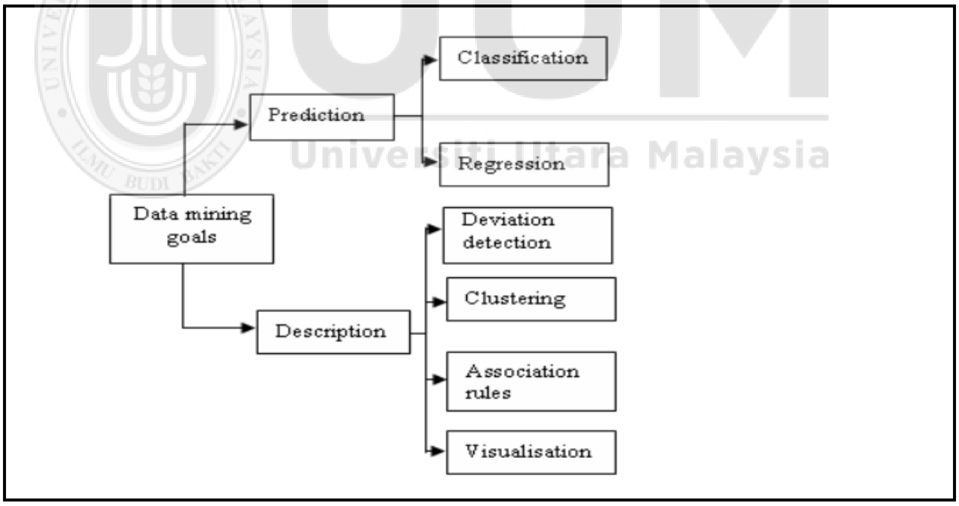

Figure 2.5. An overview of data mining methods (Danubianu, 2015)

Data mining is defined in several ways. Widely recognized as a step in a more complicated process - knowledge discovery (description) and prediction. Although this prediction is considered the primary objective of Data mining, often a description step proceeds the prediction model building. Classification or regression can achieve prediction, and deviation detection, clustering, association rules, database segmentation, sequence analysis or visualization (Danubianu, 2015). One purpose of this study is to predict learning outcomes through the eLearning system. The characteristics of the data used as a comparison of learning outcome classes. Therefore, from Figure 2.5, it can be shown that this study will use prediction methods based on classification technique.

The use of web-based eLearning systems, a large number of educational data will be generated. This extensive data generates big data in the educational sectors. Today, big data analysis techniques are used to analyze this educational data and generate different predictions and recommendations for students, teachers, and schools. This work uses collaborative filtering techniques to introduce particular courses to students based on scores received in other subjects. They are using the list of Hadoop's top Mahout Machine learning libraries to generate sets of recommendations. The similarity Log-likelihood was used to find the pattern between grades and subjects. Root Mean Square Error between actual grade and recommended grade were used to test the system. It is shown from the results of this study that schools, colleges or universities offered alternative elective courses to students (Dwivedi & Roshni, 2017).

Data mining used in educational information is intended to find useful patterns in a large number of data to optimize the study. It involves several steps. This article offers case studies for preprocessing frameworks for student outcomes prediction using data collected by Moodle (Danubianu, 2015).

Data mining offers a similar approach. But conducting deeper analysis and processing of historical data insights that can reach the point of predicting potential customers for certain service (Al Mazidi & Abusham, 2018) . Using data mining algorithms that are designed wisely to study, analyze, and examine the data interactions, data mining could provide nearly perfect anticipations of an organization's performance depending on the type of information stored (Mak, Ho, & Ting, 2011) .

Besides, according to Chien Ming, "data mining refers to the search for valuable hidden information and their correlation of a large amount of data. The related techniques include logic, statistics, and artificial intelligence. It is characterized by its thoroughness in analyzing raw data and in determining the information involved and their relations. Based on the different questions that are set, data mining techniques establish related models as the references for decision-making" (Chien Ming et al., 2007, p.18).

In addition to the details mentioned above related to data mining, there are several researchers in the field mentioned and discussed regarding data mining in eLearning. According to their study, implementation and Design of an eLearning model on the basis of Web Usage Mining (WUM) techniques, it is the research that analyzed the probable rules hidden in web log and it can help personalize the design of Web content and develop web design, customer satisfaction, and user navigation via pre-fetching and caching. An intelligent and individual platform for learners can be provided by WUM which can be considered as the most popular techniques in Web data mining (Xu & Jun, 2009).

Likewise, Han and Kamber (2000, p.7-8) have defined that "data mining is the process of discovering interesting structures from huge amounts of data where the data can be kept in databases, data warehouses, or other information repositories. It is a new interdisciplinary field, illustrating from areas such as database system, data warehousing, statistics, machine learning, data visualization, information retrieval and high-performance computing. Other technique including neural networks, pattern recognition, spatial data analysis, image databases, signal processing, and inductive logic programming".

The study of eLearning improvement is used for a lecture in the university. Then, a new developing function such as a unit for learning mode to monitor and analyze learner's learning status and a unit for searching content and analyzing content's status is explained. Furthermore, this system indicates next suitable content with data mining of learner's status and content's status by genetic algorithm (GA) after learning content. The learner could be supported by this function to maintain eLearning with well-understanding of contents and highly-motivation for learning (Ninomiya et al., 2007). Also, Yongquan, (2007) focused their study on the data mining process for discovering an interest of the user from reading behavior in the eLearning system.

Jili, Kebin, Feng, and Huixia (2009) are one group of a researcher who proposes the newly developed method to analyze behavior in eLearning by the fuzzy clustering algorithm. Their study could show that this method can reach proper implementation of eLearning behavior analyzing. Also, the teacher could understand the students' interests, personality, and other information. Cluster analysis is more effective than traditional statistical analysis methods.

Furthermore, modeling online user behavior (Milani, Jasso, & Suriani, 2008b) demonstrates the methodology for studying the relationship between activity and used time of student in using the eLearning system by used Timed Transition Automaton (TTA) algorithm for representing the output model. The finding of this study reveals that each student behavior shows different activities timelines rather than describing its pattern.

An exploratory data analysis and web log techniques are applied by users to log in to log files from Media Usage and User Session files to improve eLearning in the university Many useful usage indicators can be extracted by standardizing all access hit with the number of unique user hits. Moreover, the study found out users' interesting clustered groups and relevant rules for media access which are hidden in the system (Nukoolkit, Chansripiboon, & Sopitsirikul, 2011).

Web Usage Mining offers a variety of methods and tools for data mining to detect patterns of web usage behavior and to draw valuable knowledge for analyzing and improving web-based platforms (Dragoş et al., 2017).

Moreover, the algorithms in data mining are employed to find out patterns which identify learners both across session or groups up to their learning technique selection and purpose orientation (Mingming, 2010). Classification is one of many techniques in data mining that aims to discover the principle classification from a pre-classified data set (training data). The most proper data sets are used for databases, and a tuple represents a sample. One attribute is seen as a target attribute for classification (output), and other attributes are seen as data pattern (inputs) (Mon Fong, Shian Shyong, & Shan Yi, 1999).

Not only the above study but also Ribeiro and Cardoso (2008) conducted their study in eLearning system evaluation which has become an outstanding issue. The research aims to compare many algorithms for suitable studying in eLearning usage. Accordingly, it can be considered data mining as the process to take out knowledge from the formats in an engineering system. Therefore, the algorithms become possible productive tools to extract knowledge and model mining. It is found that "neural network, rule-associating mining, clustering and kernel-based learning methods, e.g. Support Vector Machines and Relevance Vector Machines are amongst the most used techniques for pattern classification and knowledge discovery". There are also many tools that can be used in the data mining process to investigate the suitable result for each condition.

Above all, "the support vector machine (SVM) is a controlled learning method that brings about input-output mapping functions from a group of labeled training data under a classification technique. It is a supervised learning approach the mapping function can be either a classification function, for instance, the category of the input data, or a regression function. For classification, nonlinear kernel functions are often applied to change input data to a high-dimensional feature space in which the input data turn out to be more separable compared to the original input space. Maximummargin hyperplanes are then created. The model thus produced depends on only a subset of the training data near the class boundaries" (Lipo, 2005). Thereby, the web log processing as shown in Table 2.5 which applies a support vector machine (SVM) for evaluation is also the method for this study.

Extensively, Support Vector Machines (SVM) is the algorithm in classification technique as Ribeiro and Cardoso presented that data mining is mostly the draw out process of knowledge from engineering system patterns. Algorithms in conceptual learning provide potential productive tools for extracting knowledge and pattern mining. Neural networks, rule associating mining, clustering and kernel-based learning methods (Support Vector Machines and Relevance Vector Machines) are amongst the most used techniques for pattern classification and knowledge discovery. Their studies focus mainly on the most used learning models, in particular, neural networks and SVM (Ribeiro & Cardoso, 2008). They also compared the average accuracy between seven algorithms for web log analysis, the highest rate among others data mining algorithm is SVM as shown in Table 2.6.

Table 2.6

Comparison of Data Mining Algorithm (Ribeiro & Cardoso, 2008)

	Learning Classifiers						
Algorithm	**NAIVE BAYES**	**RANDOM FOREST**	**ADABO OST**	**MLP**	**RBF**	**ENSE MBLE**	**SVM**
Average Accuracy	80.37	81.59	80.75	79.02	79.75	79.01	82.18

This topic explores the suitable data mining algorithm which will be utilized in this research. Due to Table 2.6, SVM is the most suitable algorithm for employing data mining in their studying. However, this research will use a variety of algorithms to prove that SVM will be the most suitable as this case study. Perhaps there are other interesting algorithms as well such as ZeroR, Naïve Bayes, J48, Dicision Table and Random Tree.

2.10 Measuring Learning

Chance (2003, p.36-56) also makes claims about the issue of learning to measure that "learning can be measured as a change of error, topography, intensity, speed, latency, or rate of behavior. There are also other ways of measuring learning as well. The point is that they cannot study learning unless they can measure it in some precise way". However, there is more to study learning on measuring it that could be a limitation of the research design.

Before measuring the usability of the eLearning applications, users are faced with the unknown risk of failure of applications. This article lists five metrics to measure the usability of the eLearning system through the user interface. The metrics are the time of user feedback, the average of using help methods, the average of using undo, average time spent in any page, and average use of eLearning learning system search engines. They focus on measuring the use of e-systems, not content (Elfaki et al., 2013). The proposed metrics can help to understand and evaluate the degree of end-user acceptance. Besides, the learning may be studied in many different ways. Anecdotal and case study evidence are unreliable, though good sources of hypotheses. Descriptive studies can provide useful and reliable information but cannot account for why a phenomenon occurs. Because of these limitations, learning is usually studied employing experiments; experiments allow us to see the effects of independent variables on dependent variables.

According to a new interactive eLearning system including hardware construction and software implementation (Eryou & Jun, 2008) questions can be asked under the class content and students can electronic show of hands or try to answer first using the buttons on the desk. The students can be chosen to answer the questions and offer comments to the students. The answers and the comments are displayed on the screen and kept in the database. The results of this performance can help giving scores in class which can be able to enhance students' achievement rather than real interactive teaching and learning in a classroom.

Moreover, their study can also show that teachers could control the study status of students and classroom, which is essential to raise the quality of teaching. Furthermore, the interaction between teachers and students and the competition among students in class helps to make the class more interesting, challenging and worthy and the previous passive learners are adapted to be active participates. So, it also showed the increasing level of the eagerness of learning the eagerness of learning which can represent the teaching success considerably.

An active learning study found the students' requirements in learning types that preferred from the most to the lease as experienced learning, role-playing learning, case-study learning, discussion learning, assignments/homework and lecture learning consequently. This result is suitable for designing the workflow and functional modules of the active eLearning (Chun Xia et al., 2010).

2.11 Theoretical Framework

Researches on the learning theory of eLearning mentioned on the relevant theories such as learning theory, teaching theory, behaviorist learning theory, the development of learning and teaching theory, instructional design theory, concept of eLearning system design, eLearning teaching objective related to teaching objective for promote learners to study (Chen, Wu, Song, & Chen, 2009; Gang, 2010; L. Wu, Xu, & Qu, 2009). These theories could be the basis for the research's theoretical framework to demonstrate the research design as depicted in Figure 2.6.

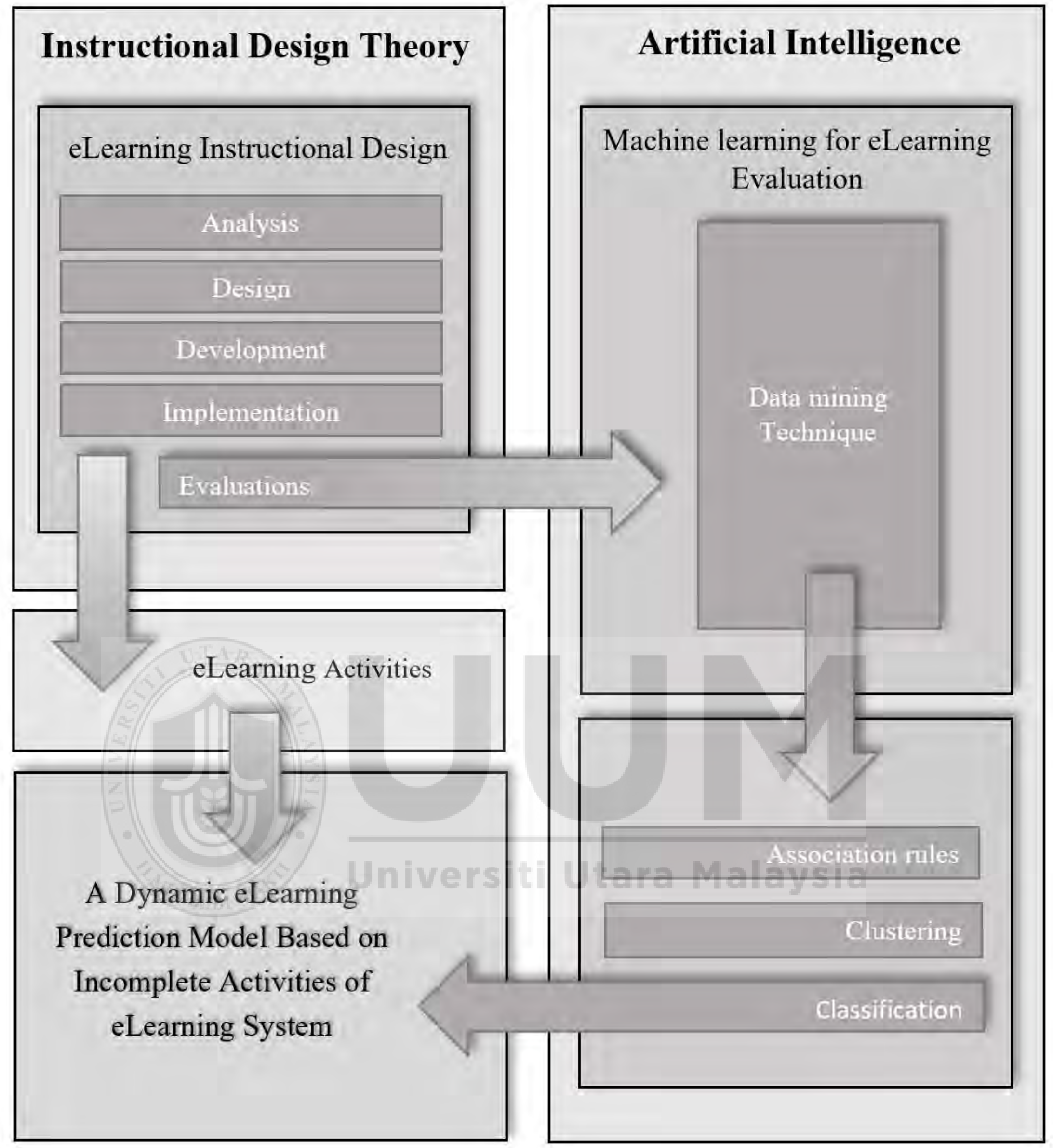

Figure 2.6. Theoretical framework: A dynamic eLearning prediction model based on incomplete activities of eLearning system

According to this theoretical framework (Figure 2.6), eLearning evaluation which is a part of the eLearning instructional design (Khalil & Elkhider, 2016) applied on data mining technique for eLearning evaluation process to meet eLearning teaching objectives. For creating eLearning usage model, favorably three main algorithms are considered which are clustering, classification as well as association rules. For classification problems, naturally, it aims to measure a classifier's performance regarding the error rate. The classifier predicts the class of each instance: if it is correct, that is counted as a success. Otherwise, it is an error. The error rate is only the proportion of errors made over a whole set of instances, and it evaluates the whole performance of the classifier (Witten & Frake, 2005). In the same way, this study will use a different algorithm for classifying the eLearning usage that could be able to predict learner achievement to meet the unknown activities relationship inside eLearning system based on incomplete eLearning system. eLearning activities are the essential source for data analysis process.

2.12 Summary

eLearning is a tool that supports the latest learning systems. eLearning means that anytime, anywhere learning and education are available through computer technology. The purpose of the learning system is to develop learning quality. Different learning styles will improve learner's performance. Automation of the pedagogical approaches to deliver learning processes remotely, enabling interaction in a learning environment and attaching to students consistently with learning processes are the primary objective of eLearning.

Many educational institutions are committed to improving the quality of their education and their students. Learning outcomes are the response to such intentions, predicting student, a performance by analyzing their learning behavior is one of the best chance to take into account. When predicting performance, it will be easy for teachers, school authority or other related parties to determine the appropriate policies on the issue. eLearning courses are very demanding in recent times. The need to study student achievement and predict their performance will also increase. With the growing popularity of educational technology, data mining algorithms that are well suited for predicting student performance have been reviewed.

In conclusion, many interesting topics are supporting this work. To facilitate students' adoption and use the eLearning service channel, educational institutions and eLearning researchers need to understand the activities that influence students' adoption and usage of eLearning recommender systems. Many activities should be considered when integrating the course content into the eLearning format, which is mostly under the context of the curriculum and educational platform to be published. Student performance is an essential measure for evaluating the effectiveness of the eLearning platform. However, the problem arises that it is not unique for evaluating the quality of eLearning methods; other activities also play an essential role. Performance assessments in the context of most eLearning are conducted using online tools to examine the effectiveness of learners according to their knowledge. The previous studies in eLearning usage have many purposes, for instance, to predict the eLearning usage to high achievement. Most of these purposes were studies based on data mining techniques that consist of sub-techniques and algorithms.

Based on this literature review, there are guidelines such as the eLearning activities studying, the learning model developing, and model effectiveness evaluating. These guidelines provided in this chapter review will be used to design the research methodology in the next chapter.

CHAPTER THREE
METHODOLOGY

3.1 Introduction

The primary goal of this research is to develop an appropriate eLearning outcome prediction model for accomplishing effective educational development. The incomplete activities related to eLearning usage are hidden in the users' activities history called web log. This web log is essential as a data source for processing data mining techniques to find appropriate models. Therefore, the study of the web log could assist the developers in developing more stable, productive and predictable eLearning based on the usage models.

According to the problem statement of this study, the eLearning activities affected by high learning outcome prediction on incomplete eLearning system is the main issue for more understanding.. Thereby, this study needs to examine the following research methodology.

3.2 Research Design

The research design of this study could be demonstrated by dividing the process into seven steps as shown in Figure 3.1.

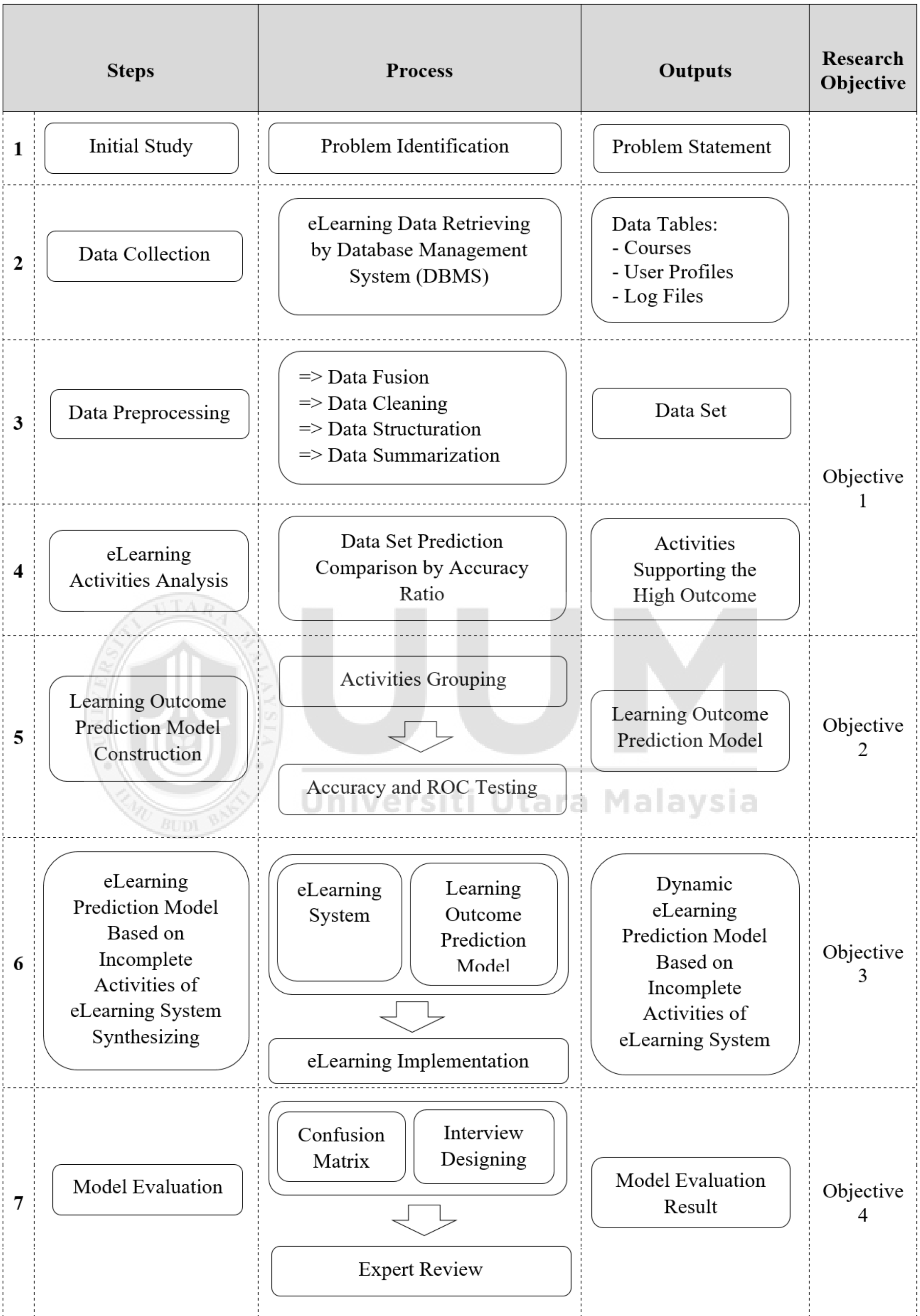

Figure 3.1. Research Design

Step one is initial study that gathered the current eLearning system usage status and its problem. Then identify the problem based on the real-world usage from previous research and in actual eLearning implementation.

Step two is the data collection process that retrieves the eLearning data from the database management system (DBMS) software name MySQL. The data are the tables of student profiles and student grade results from university student database. Other includes the eLearning system tables, which are learning resources tables, learning activities tables and learning quizzes tables.

Step three is data preprocessing that generates data set by four steps - data fusion, data cleaning, data structuration, and data summarization. This outcome of this step is the dataset that consist of eLearning activities suitable for analysis process.

Step four is the eLearning activities analysis that analyses the activities to determine whether it passes the accuracy ratio's acceptable value or not. The activities accepted by accuracy ratio standard level will be the activities to support great outcome prediction.

Step five is learning outcome prediction model construction. This step is an iterative work to group the activities by factor analysis technique or grouping of elements by using similar names. Many predictive algorithms are tested until the accuracy ratio is reached. At the same time, the ROC values obtained from the confusion matrix procedure are used as a basis for consideration. If the level of prediction by the accuracy ratio is high (> 0.75%) and the model efficiency (ROC) is higher than low discrimination level (> 0.5%), this model is chosen as a prototype for further use.

Step six is eLearning prediction model based on incomplete activities of eLearning system synthesizing. This process synthesize the learning outcome prediction model with the eLearning system that will affect the implementation of eLearning in higher education institutions. While the output of the learning outcome prediction model is still able to find the appropriate activities for the prediction, these output activities may be the suitable activities which can also have acceptable predictive effects. Therefore, this whole model can be classified as a dynamic model, which is applicable in many situations.

The last step is an expert review of the advantages, accuracy and effectiveness for the dynamic eLearning prediction model based on incomplete activities of eLearning system by experts. This expert review focuses on the level of consensus, as well as suggestions for improvements to make the model better.

3.3 Data Collection

For eLearning research, the student's usage have been collecting and analyzing web usage data obtained by the course administrators, which is called web log mining. The periodic statistics generated by web log mining show patterns of website usage. The patterns show the eLearning usage pattern and needs (Chanchary, Haque, & Khalid, 2008; Vivekananthamoorthy, Sankar, Siva, & Sharmila, 2009).

This study aims to collect the comprehensive data and can be used for creating sufficient models from two groups of data. Two groups of data are collected which are eLearning system web log and student learning profiles that recorded the student learning activities as students' usage history and students' learning database are shown in Figure 3.2. These two parts of data will be used for extracting and learning result respectively for this research methodology.

An eLearning system web logs are the tables of students activities recorded that appear in eLearning system tables (Table 2.3 and Table 2.4). The data collection from web log files is exported from eLearning system. For this case, web log originates from either student or lecturer usage as designed by lecturer.

A students' learning database is the students' personal data such as name, last name, metric id, status, grade etc. The personal data files are exported directly from student's learning database in order to use in the data extraction process.

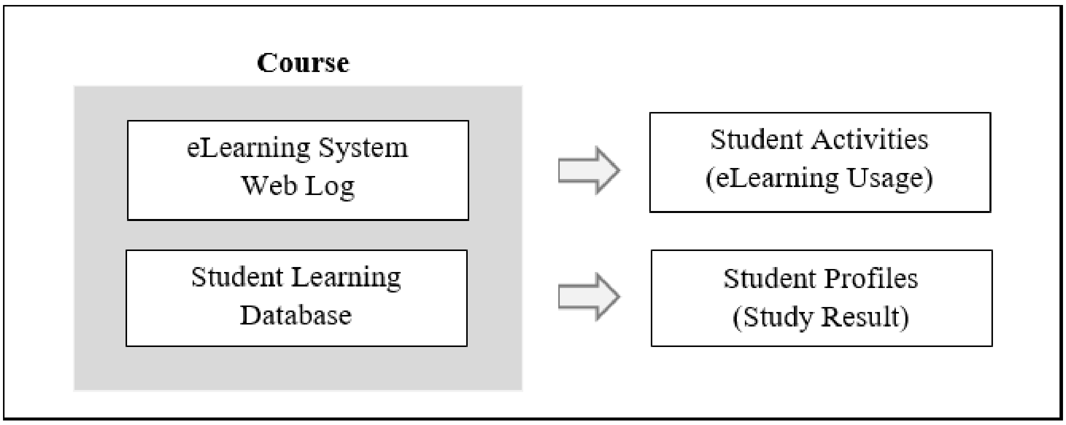

Figure 3.2. Two groups of data collection in each course

For courses sampling, the traditional courses and eLearning courses have the whole activities from the beginning to finishing point within a semester. All usage history and student results are kept in a database and web log in the same session. Thereby, web log collected from each course within one semester is sufficient for the study used. However, this study has been collecting eLearning log from six semesters during 2012-05-25 to 2015-04-06 to support more data mining technique performance and more variety of eLearning activities.

Accordingly, the purpose of this study is an approach to appropriate eLearning usage models for learners' outcome prediction. The target rate of predicting performance is determined as at least 75% (Witten & Frake, 2005). Each web log from the selected courses should consist of three activities for eLearning usage classification processing; counting, timing and scoring as shown in Table 2.5 in order to get the proper outcome. Thus, this study population is all eLearning course from six semesters during 2012-05-25 to 2015-04-06 and the sample size are eLearning courses that can pass the process of data preprocessing step.

3.4 Data Preprocessing

The data preprocessing is the step for preparing the appropriate data from web log in order to use in data analysis step. Classical data preprocessing involves three steps: data fusion, data cleaning, and data structuration. Tanasa and Trousse (2004) solution for web usage mining (WUM) adds what they call advanced data preprocessing that consists of a data summarization step, which will allow the analyst to select only the information of interest. Thereby, this study uses these four important steps as data fusion, data cleaning, data structuration, and data summarization in order to complete the data preparation step.

3.4.1 Data Fusion

Data fusion is the dataset integration process. This study combines the Moodle log data set with the student profile data set to be the new data set. For example, this data fusion process combines three tables as student profile, student grade and student activities to be the tabled as student profile-activities. The student profile is the table of student master file such as the student_ID, student_name, student_program, etc. The student grade is the table for storing the learning grade result of each subject for all students. Student activities is a table of all activities after login into eLearning. This table stores a new record after every click on every link in eLearning. The new record of student activities will keep the data such as student ID, timestamp, subject ID and activities ID.

The data fusion process will join the three tables to be the new table name student profile-activities. This table is the source table necessary to the data cleaning process as shown in Figure 3.3.

Student Profile		
Student_ ID	Student_ Grade	Student …
S001	B	…
S002	C	…
S003	D+	…
S004	A	…

Student Grade		
Student_ ID	Student_ Subject	Student_ Grade
S001	C001	B
S002	C001	C
S003	C001	D+
S004	C001	A

1 1

M M

Student Activities						
Log_ ID	Log_ Time	Student_ ID	Student_ Subject	Activity_ 01	Activity_ 02	Activity …
001	Time1	S001	C001	1		…
002	Time2	S002	C001	1		…
003	Time3	S001	C001		1	…
004	Time4	S001	C001	1		…
005	Time5	S003	C001		1	…
006	Time6	S004	C001	1		…

Student Profile-Activities							
Log_ID	Log_Time	Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity …
001	Time1	S001	C001	B	1		…
002	Time2	S002	C001	C	1		…
003	Time3	S001	C001	B		1	…
004	Time4	S001	C001	B	1		…
005	Time5	S003	C001	D+		1	…
006	Time6	S004	C001	A	1		…

Figure 3.3. Data fusion process

3.4.2 Data Cleaning

Data cleaning is the process of filling in or disposing of data to be complete and ready for use in the next step. For example, this data cleaning process replaces the blank values to the new values. The acquisition of information is possible in several ways. The first method is to add a zero to be used in comparison with other values contained in non-empty data. The zero replacement suitable for the yes/no data type such as do the activity or do not do the activity. Another way is to simulate the data to fit the existing data. Data replication is a way to be careful and have enough reason to use that information. At the same time, it necessary to use the other statistics required to complete the simulation as shown in Figure 3.4

Student Profile-Activities								
Log_ID	Log_Time	Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity_ 03	Activity_ 04
001	Time1	S001	C001	B	1			
002	Time2	S002	C001	C	1			
003	Time3	S001	C001	B		1		
004	Time4	S001	C001	B	1			
005	Time5	S003	C001	D+		1		
006	Time6	S004	C001	A	1			
007	Time7	S003	C001	D+			1	

Student Profile-Activities								
Log_ID	Log_Time	Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity_ 03	Activity_ 04
001	Time1	S001	C001	B	1	0	0	0
002	Time2	S002	C001	C	1	0	0	0
003	Time3	S001	C001	B	0	1	0	0
004	Time4	S001	C001	B	1	0	0	0
005	Time5	S003	C001	D+	0	1	0	0
006	Time6	S004	C001	A	1	0	0	0
007	Time7	S003	C001	D+	0	0	1	0

Figure 3.4. Data cleaning process

3.4.3 Data Structuration

Data structuration process is the process of improving the structure of the data table to suit the application. For example, this data structuration process reduces unnecessary fields or collapsing fields from multiple fields together. This example the field name activity_04 contain the empty value when passed the data cleaning process this field become the new value as zero. However, all record of this field represent an activity data. Then activity_04 is unnecessary for the next process. It may be concluded that, data structuration collect only field contained the data that meaningful enough to be processed and ready for use in the next step as shown in Figure 3.5.

Student Profile-Activities								
Log_ID	Log_Time	Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity_ 03	Activity_ 04
001	Time1	S001	C001	B	1	0	0	0
002	Time2	S002	C001	C	1	0	0	0
003	Time3	S001	C001	B	0	1	0	0
004	Time4	S001	C001	B	1	0	0	0
005	Time5	S003	C001	D+	0	1	0	0
006	Time6	S004	C001	A	1	0	0	0
007	Time7	S003	C001	D+	0	0	1	0

Student Profile-Activities							
Log_ID	Log_Time	Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity_ 03
001	Time1	S001	C001	B	1	0	0
002	Time2	S002	C001	C	1	0	0
003	Time3	S001	C001	B	0	1	0
004	Time4	S001	C001	B	1	0	0
005	Time5	S003	C001	D+	0	1	0
006	Time6	S004	C001	A	1	0	0
007	Time7	S003	C001	D+	0	0	1

Figure 3.5. Data structuration process

3.4.4 Data Summarization

Data summarization is a process of gathering relevant data in different ways such as combining values, frequency counting to be used in the next step as shown in Figure 3.6.

Student Profile-Activities							
Log_ID	Log_Time	Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity_ 03
001	Time1	S001	C001	B	1	0	0
002	Time2	S002	C001	C	1	0	0
003	Time3	S001	C001	B	0	1	0
004	Time4	S001	C001	B	1	0	0
005	Time5	S003	C001	D+	0	1	0
006	Time6	S004	C001	A	1	0	0
007	Time7	S003	C001	D+	0	0	1

Student Activities-Log						
Student_ID	Student_ Subject	Student_ Grade	Activity_ 01	Activity_ 02	Activity_ 03	Activity …
S001	C001	B	2	1	0	…
S002	C001	C	1	0	0	…
S003	C001	D+	0	1	1	…
S004	C001	A	1	0	0	…

Figure 3.6. Data summarization process

For Figure 3.6, data summarization process prepares the information to be ready for the data mining process. There are several ways to prepare such information, depending on the purpose of the data that will lead to the processing. In this example, multiple records with duplicate student identifiers are collapsed and the values of the numbers in each field are added together. In this case, it is the sum of the numbers representing the same activity frequency of the same course to explain how many students did one of the activities in the course. At the same time, it also included the frequency of other activities together. The final result of this process is to represent each student's record that identifies the frequency of each activity in a given subject. This result is intended to guide the learning process of learners in the next step. Sometimes, the use of data in different fields may be scaled to zero-one (0-1), depending on the needs and goals of the user.

3.5 eLearning Activities Analysis

The eLearning usage model helps eLearning users that include learners and teachers for evaluating and predicting results of learning beforehand. Therefore, the eLearning usage model could be improved positively which can affect learners' higher performance. Unfortunately, some higher institution researchers use their web logs for developing appropriate and effective model but only for particular uses. For instance, the study of eLearning usage evaluation that can enabled the new model which related between eLearning usage and effect, and can be used to evaluate students in their eLearning system (Lingyan et al., 2010).

The data collected from course samples are analyzed using data mining technique, in particular ZeroR, Native Bayes, SVM, J48, Decision Table and Random Tree algorithms according to the following steps.

The first step is data preparation. There are four tasks for processing in this step, which consists of data fusion, data cleaning, data structuration, and data summarization. As well the attributes in web log will be classified into three groups as counting activities, timing (period) activities and scoring activities (Lingyan et al., 2010) (see also in Table 2.5).

The second step is data extraction. This study wanted to compare the results of the prediction with the grade level of learning that was categorized explicitly. Hence, this step will be processed by classification technique. This model aims at predicting learners' grading results that can be effective to learners' performance improvement.

The data obtained from the first step is divided into two groups; training data and test data. Then, a model will be constructed by various algorithms with training and test data up until an accuracy classification model ensures it. Normally, researchers intend to use the model for prediction of future outputs from inputs. A model's ability to accurately predict outputs for future inputs is called generalization. In order to determine how well the model will generalize to future data generated by the same process, we can hold out some of the data that we used to fit the model and use it to test the model. The set that we hold out call the test set, and the set we use to train the model call the training set. A commonly used split it to use 80% of the data for the training set, and the remaining 20% for the test set.

For this study, the partition of data for training and test data set is divided by K-fold cross-validation method, which is practically the standard method. Tests have also shown that the use of stratification improves results slightly. Thus, the standard evaluation technique in situations where only limited data is available is stratified K-folds cross-validation. However, the quantity of the partition in this study can be increased up to the accurate error estimation, which target of predicting performance will determine the success rate at least 75% (Witten & Frake, 2005). All data analysis is shown in Figure 3.7.

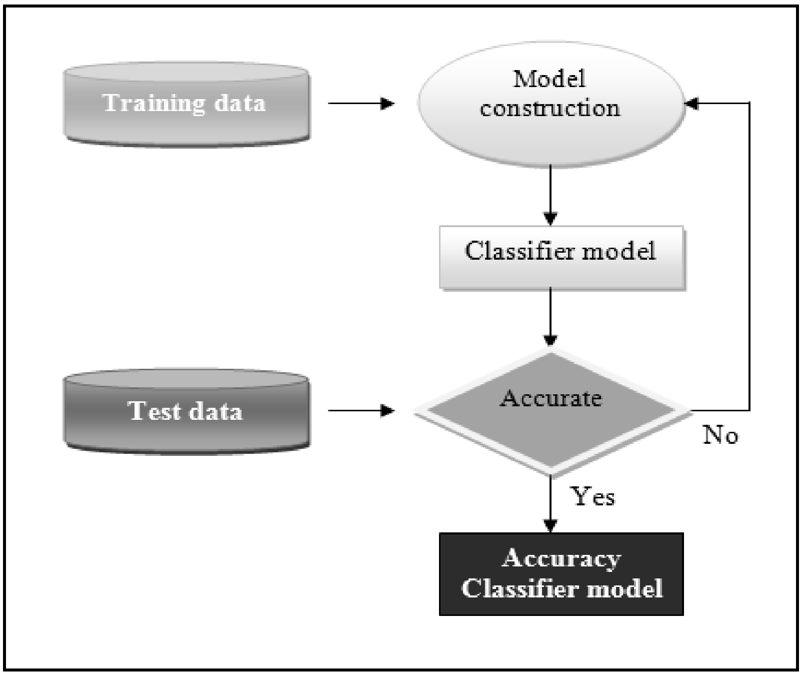

Figure 3.7. Accuracy classifier model finding

Figure 3.7 presented that the predictive modeling of eLearning system log has a circular flow of modeling. Using algorithms to determine accuracy, if the value of accuracy does not meet the requirement of 75% or more, then go into the process of preparing the data and continue to find the accuracy. In this task, we use software called WEKA (Waikato Environment for Knowledge Analysis) as the main tool for management. It has the ability to handle log data and at the same time it can find the required precision through the use of many available algorithms. From WEKA, there are algorithms group in classification menu. This work chooses the representative algorithm from each menu group based on the recommendations from the previous eLearning research and considers the results of the algorithm that higher than other algorithms.

3.6 eLearning Outcome Prediction Model Construction

From the previous topic, it is an analysis of the activities occurring in the eLearning system to find out whether the activities have sufficient information to be used in predicting the learning outcomes. From this process, a number of courses are ready for use in the process of predicting learning outcomes in the next step. The concept of this research does not focus on any particular activities. It will focus on the activities that have the potential to predict the learning outcomes. This variant has a tremendous variety of activities because of the use of all activities in the processing of the appropriate accuracy analysis.

Therefore, the repetition of these activities for the purpose of studying the relationship of those activities is even more profound. In general, finding the relationship of a group of activities in a basic statistical way is often a factor analysis technique. At this stage, factor analysis technique is used as the first choice for processing. Considering the initial activities, we know that there are some similarly named activities, because one of the activities that occurs on eLearning system is the manipulation of several activities. For this reason, there is a second alternative to distinguish these activities by using similar names for grouping. From these two alternatives, it is the process of finding a correlation of all the factors derived from the previous step.

From all of the above steps, the processing sequence of the predictive learning process is processed by the accuracy classifier model finding and followed by grouping the activities with the two-grouping type discussed.

After the segmentation step, there will be a group of activities that are likely to be used as models for predicting learning outcomes. Measurement of accuracy ratio and ROC values will be the final step to determine which groups of activities have the potential to be used to predict learning outcomes.

By all of the mentioned sequences, it can be shown as a modeling step to find the appropriate activities for predicting learning outcomes, as shown in Figure 3.8.

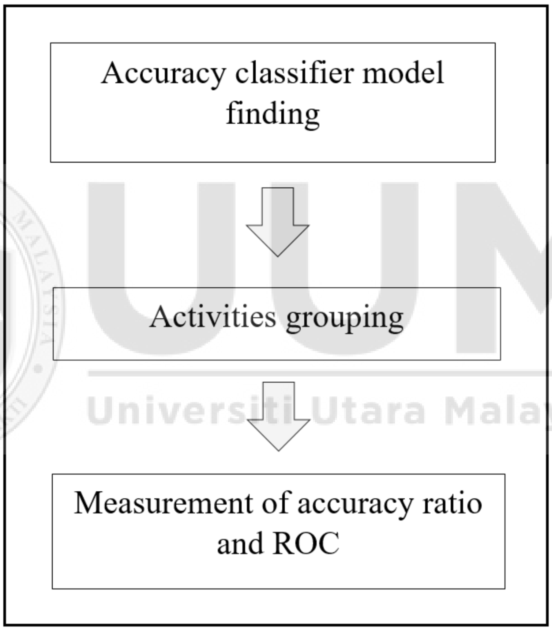

Figure 3.8. eLearning outcome prediction model construction

3.7 Model Synthesizing

This research focuses on enhancing the potential of the existing eLearning systems of educational institutions. By studying in a classroom without storing information, activities cannot be used to analyze data to find out what is going on in the future. On the other hand, academics who have adopted the eLearning system are already using it. If there is a research concept that does not give sufficient research results, the resulting prediction tool cannot be used consistently. In this phenomenon may be said that the concept is not good enough to reduce both the time and cost of finding tools for predicting good learning outcomes. In this research, we have the idea to extend that research by creating tools for finding models for predicting good learning outcomes, focusing on the elasticity of the model to be obtained. Therefore, this model design will present the search and use the model for predicting the results simultaneously to provide an overview of all the goals of the research.

Most eLearning systems and Moodle have the ability to develop plug-ins. The eLearning plug-in is for predicting grades before the end of the semester. This plug-in will come from the learning outcome prediction model. Creating a learning outcome prediction model is a process that can find the appropriate activities for predicting the outcome. In the process of model synthesis, it is necessary to present the relationship and procedure to apply the learning outcome prediction model to the educational management system of higher education institutions which are using the eLearning system. The concept of applying eLearning in education to develop learners' potential is the basis for all educational institutions. If an eLearning system is more effective than traditional classroom learning, online investment will be more rewarding.

The above-mentioned related components and previous research reviewing will be combined to provide an overview model of the learning outcome prediction model to the existing eLearning system implementation, which can dynamically predict the learning result to meet the high learning outcomes.

In the synthesizing sequences, there must be activities that can be credible that the emerging model can explain all relationships of interest. The processing step collects all the relevant elements first, and then analyzes the relationship of each element that is associated with the other elements in any direction. In the final stage, the correlation between all elements is drawn on the basis of the analysis. The steps of modeling can be seen in Figure 3.9.

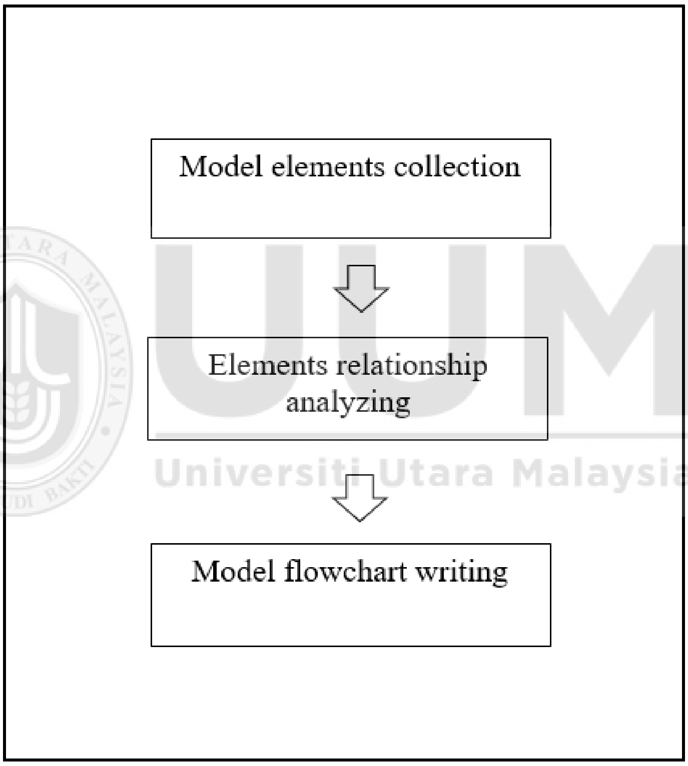

Figure 3.9. eLearning model synthesizing

As shown in Figure 3.9., the result is modeled name as dynamic eLearning prediction model based on incomplete activities of eLearning system.

3.8 Model Evaluation

This study has created an overview model and sub-model. To confirm that these models are reliable, this work will use both machine learning measurement and expert assessment. For machine learning measurement will use accuracy and confusion matrix process and for expert assessment will use the expert review process as follow.

3.8.1 Accuracy and Confusion Matrix Process

A confusion matrix (Kohavi and Provost, 1998) contains information about actual and predicted classifications done by a classification system. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two-class classifier. The entries in the confusion matrix have the following meaning in the context of this study:

"a" is the number of correct predictions that an instance is negative,

"b" is the number of incorrect predictions that an instance is positive,

"c" is the number of incorrect of predictions that an instance negative, and

"d" is the number of correct predictions that an instance is positive.

Confusion matrix for a two-class classification shown as Figure 3.10.

		Predicted class	
		Negatives	Positives
Actual class	Negatives	a	b
	Positives	c	d

Figure 3.10. Confusion matrix for two-class classification

The accuracy (AC) is the proportion of the total number of predictions that were correct. It is determined using the equation: AC = (a+d)/(a+b+c+d)

The recall or true positive rate (TP) is the proportion of positive cases that were correctly identified, as calculated using the equation: TP=d/(c+d)

The false positive rate (FP) is the proportion of negative cases that were incorrectly classified as positive, as calculated using the equation: FP=b/a+b

The true negative rate (TN) is defined as the proportion of negative cases that were classified correctly, as calculated using the equation: TN=a/a+b

The false negative rate (FN) is the proportion of positive cases that were incorrectly classified as negative, as calculated using the equation: FN=c/c+d

Finally, precision (P) is the proportion of the predicted positive cases that were correct, as calculated using the equation: P=d/b+d

Receiver Operating Characteristics (ROC) is relationship between true positive rate (TP Rate) and false positive rate (FP Rate).

3.8.2 Expert Review Process

The expert review process was assessed by eight experts. The chosen criteria are the experience on eLearning teaching, eLearning administration and eLearning data mining. The main content of the expert review process is the evaluation form that consists of:

- Personal Profile Information.
- Model details.
- Review information on each issues.
- The details of model questions issue for the expert assessment include:
    - The accuracy of the learning outcome prediction model, which is a sub-model by the accuracy ratio and ROC result.
    - The main model advantage.
    - The main model effectiveness.

In addition, the experts discussion about defects of the model should be filled (e.g. model development, performance measurement, model impact study). The information obtained from the experts is divided into descriptive data and numerical data. The processing of the numerical data will use Likert scale methods. Likert scale have been used to measure character and personality traits. The difficulty of measuring attitudes, character, and personality traits lies in the procedure for transferring these qualities into a quantitative measure for data analysis purposes (Harry N. Boone & Boone, 2012). For the descriptive data, it will use the summarize processing to analyze each issue in question.

3.9 Summary

This research methodology consists of seven stages as initial study, data collection, data preprocessing, eLearning activities analysis, learning outcome prediction model construction, a dynamic eLearning prediction model based on incomplete activities of eLearning system analysis and model evaluation. The data collection part is the courses sampling for web log collection. The samples of this study are from every course taken from eLearning system case study, which is completed with processing components. The selected institution is from a university in Thailand, which has been using the eLearning system (Moodle). The university selected for this study is Suan Dusit University, Thailand. eLearning usage model extracting will be used data mining technique for eLearning usage in order to generate an eLearning usage model. Analyzing process by the factor analysis to approach the relationship of eLearning activities that affect high learning outcome prediction. Finally, the model evaluation process is used to build credibility from relevant experts.

CHAPTER FOUR
IMPLEMENTATION AND RESULT

4.1 Introduction

The main purpose of this research is learning how to use the eLearning to predict students' achievement based on eLearning incomplete activities. The ability to predict the behavior of different learners could be applied to the eLearning system in order to assist both the learner and the instructor. This research also develops a model for the university's teaching and learning development as well as for those departments responsible for education management by using modern technology. In this work, the different approaches to reach a learning outcome prediction model by machine learning technique is more suitable for predicting tools. This case study works on Moodle eLearning system. Its open source system, which is suitable for the developer to share the experiences. This work studied on eLearning system to collect log data for studying the student behavior in order to predict the learning outcomes before the end of the course. The data and results of this finding will depend on their context.

4.2 eLearning Data Characteristic

This study has collected log data from eLearning database from 2012-05-25 to 2015-04-06. These three years' log takes six semesters study period. There are 53 courses provide by eLearning. After log cleaning, processing and matching to the available grade result there are 20 courses in total. The chosen courses were used by courses users and stored the activities log number as shown in Table 4.1

Table 4.1

Chosen eLearning Courses from Log

Course ID	Log number (record)	Student Number
101	571	22
107	533	23
12	620	20
126	4306	53
127	3063	28
128	1243	18
146	686	17
149	5190	45
165	11991	50
178	2065	48
190	1680	25
191	2691	30
195	1386	18
204	4272	13
206	1061	24
28	508	15
29	789	43
33	4078	30
34	993	31
36	948	13
Total	48674	566

Table 4.1 presents the result of preparing and cleaning process data by data mining technique. The total number of courses is 20, the maximum number of logs per course is 11,991 and the minimum number of logs per course is 508 records. The maximum number of students per course is 53 and the minimum is 13. The courses selected in this study are available courses during 2012-05-25 to 2015-04-06 in which various activities were recorded.

4.3 eLearning Activities

The Moodle eLearning is the famous platform that transforms the real-world classroom into the virtual classroom. There are many activities in this eLearning system for the user. All eLearning activities designed by its activities combination. Thus, there is a large number of activities have been used in this system. For this case study eLearning system was found 102 possible activities as shown in Table 4.2

Table 4.2

eLearning Possible Activities

Activities ID	Activity Name	Activities ID	Activity Name
1	admin_report capability	52	forum_view discussion
2	assignment_add	53	forum_view forum
3	assignment_update	54	forum_view forums
4	grades	55	forum_view subscribers
5	assignment_upload	56	label_add
6	assignment_view	57	label_update
7	submission	58	lesson_add
8	blog_view	59	library_mailer
9	calendar_add	60	login_error
10	calendar_edit	61	message_add contact
11	chat_add	62	message_block contact
12	chat_report	63	message_remove contact
13	chat_talk	64	message_write
14	chat_view	65	notes_view
15	choice_add	66	page_add
16	choice_choose	67	page_update
17	choice_update	68	page_view
18	choice_view	69	quiz_add
19	course_add mod	70	quiz_addcategory
20	course_delete	71	quiz_attempt
21	course_delete mod	72	quiz_close attempt
22	course_editsection	73	quiz_continue attempt

Table 4.2 (Continue)

23	course_enrol	74	quiz_delete attempt
24	course_new	75	quiz_editquestions
25	course_recent	76	quiz_manualgrade
26	course_report live	77	quiz_manualgrading
27	course_report log	78	quiz_preview
28	course_report outline	79	quiz_report
29	course_report participation	80	quiz_review
30	course_report stats	81	quiz_update
31	course_unenrol	82	quiz_view
32	course_update	83	quiz_view all
33	course_update mod	84	quiz_view summary
34	course_view	85	resource_add
35	discussion_mark read	86	resource_update
36	folder_add	87	resource_view
37	folder_edit	88	role_assign
38	folder_update	89	role_edit
39	folder_view	90	role_override
40	forum_add	91	role_unassign
41	forum_add discussion	92	url_add
42	forum_add post	93	url_update
43	forum_delete discussion	94	url_view
44	forum_delete post	95	user_change password
45	forum_search	96	user_login
46	forum_subscribe	97	user_logout
47	forum_subscribeall	98	user_update
48	forum_unsubscribe	99	user_view
49	forum_update	100	user_view all
50	forum_update post	101	workshop_add
51	forum_user report	102	workshop_view

Table 4.2 shows 102 possible activities. These activities come from log checking that is available in the log table. eLearning activities provided the user's action for students and teachers click. Each action became the data and stored in the system by related module processing. After users are done the activities the log record have stored in the log table at the same time.

The log-cleaning process is the frequency counting on each users action and on each activities grouping by data mining task as described in data preprocessing part from chapter three, there were 20 activities remaining that can proceed in the prediction process as shown in Table 4.3

Table 4.3

eLearning activities after Cleaning Process

Activities ID	Activity Name	Activities ID	Activity Name
1	quiz_view	11	forum_view_discussion
2	quiz_view_summary	12	assignment_upload
3	quiz_continue_attempt	13	forum_add_post
4	quiz_close_attempt	14	forum_view_forum
5	quiz_attempt	15	forum_delete_discussion
6	quiz_review	16	forum_add_discussion
7	url_view	17	forum_delete_post
8	course_view	18	forum_unsubscribe
9	assignment_view	19	forum_subscribe
10	resource_view	20	forum_update_post

4.4 Data Set Classification Scheme

From Table 4.3 shows the 20 activities after cleaning process that will be used in classification process. These activities are the source of different student grade result in each course. In the classification process of data mining one needs to prepare a suitable data set for research objective. One of the study targets is to classify the learning result for the best prediction. Therefore, this study separated grade result into 3 types as shown in Table 4.4

Table 4.4

Grade Result Scheme for the Best Prediction

Data Set Class Scheme	**Good Learning Result**	**Average Learning Result**	**Poor Learning Result**
Eight classes data set		A,B+,B,C+,C,D+,D,F	
Three classes data set	A,B+,B	C+,C	D+,D,F
Two classes data set	A,B+,B,C+,C	-	D+,D,F

*Note:
Grade point (%): A >= 90, B+ >= 85, B >=75, C+ >=70, C >=60, D+ >=55, D >=50, F < 50

Table 4.4 shows 3 types of grade result grouping. Eight classes data set is for the prediction process based on 8 groups classification A, B+, B, C+, C, D+, D, F. Three classes data set is for the prediction process based on the 3 groups of classification with A, B+, B as the good learning group, C+, C as the average learning group and D+, D, F as the poor learning group. Two classes data set is for the prediction process based on 2 groups of classification with A, B+, B, C+, C as the good learning group and D+, D, F as the poor learning group.

After the three data sets were prepared, a classification process was run by several algorithms. The algorithms chosen was the standard option.

The eight classes data set processed by the classification algorithm is shown the result in Table 4.5

Table 4.5

Eight Classes Data Set Classification Accuracy Ratio (%)

Course ID	**ZeroR**	**Naive Bayes**	**SVM**	**J48**	**Decision Table**	**Random Tree**	**Average by Course**
101	20	25	30	30	30	15	25.00
107	39.13	17.39	34.78	47.82	39.13	17.39	32.61
12	30	20	10	25	20	30	22.50
126	56.6	50.94	64.15	54.71	66.03	58	58.41
127	32.14	21.42	28.57	14.28	32.14	14.28	23.81
128	33.33	5.5	33.33	33.33	33.33	27.77	27.77
146	52.94	23.52	52.94	35.29	52.94	47.05	44.11
149	31.11	24.44	17.77	46.66	40	37.77	32.96
165	24	34	34	28	32	26	29.67
178	31.25	20.83	25	25	31.25	27	26.72
190	36	28	32	40	36	32	34.00
191	50	43.33	50	40	53.33	46.66	47.22
195	50	38.88	50	44.44	50	11.11	40.74
204	0	23.07	23.07	53.84	7.69	15.38	20.51
206	45.83	45.83	50	29.16	45.83	33.33	41.66
28	40	33.33	40	26.66	26.66	26.66	32.22
29	34.88	62.79	37.2	83.72	79.06	76.74	62.40
33	30	26.66	33.33	20	30	16.66	26.11
34	22.58	38.7	29.03	83.87	64.51	61.29	50.00
36	61.53	38.46	61.53	38.46	61.53	46.15	51.28
Average by Algorithm	36.07	31.10	36.84	40.01	41.57	33.31	36.48
All Course	25.4	18.99	26.2	23.91	25.4	21.51	23.57

Table 4.6 shows the three classes data set processed by classification algorithm.

Table 4.6

Three Classes Data Set Classification Accuracy Ratio (%)

Course ID	ZeroR	Naive Bayes	SVM	J48	Decision Table	Random Tree	Average by Course
101	68.18	86.36	59.09	77.27	86.36	81.81	76.51
107	47.82	52.17	39.13	47.82	47.82	34.78	44.92
12	55.00	45.00	50.00	45.00	55.00	40.00	48.33
126	75.41	64.15	73.58	66.03	75.47	58.49	68.86
127	50.00	53.57	50.00	46.42	42.85	46.42	48.21
128	61.11	27.77	61.11	44.44	61.11	44.44	50.00
146	88.23	88.23	88.23	88.23	88.23	76.47	86.27
149	42.22	55.55	44.44	53.33	37.77	42.22	45.92
165	40.00	44.00	52.00	54.00	50.00	44.00	47.33
178	56.25	45.83	54.16	60.41	52.08	54.16	53.82
190	68.00	48.00	68.00	68.00	64.00	72.00	64.67
191	56.66	53.33	53.33	66.66	60.00	63.33	58.89
195	66.66	50.00	66.66	66.66	66.66	50.00	61.11
204	61.53	30.76	61.53	69.23	61.53	46.15	55.12
206	79.16	66.66	79.16	75.00	79.16	75.00	75.69
28	33.33	53.33	53.33	66.66	40.00	46.66	48.89
29	58.13	74.41	58.13	100	90.69	83.72	77.51
33	33.33	50.00	50.00	63.33	36.66	56.66	48.33
34	51.61	70.96	54.83	93.54	93.54	96.77	76.88
36	69.23	38.46	69.23	69.23	69.23	61.53	62.82
Average by Algorithm	58.093	54.927	59.297	66.063	62.908	58.73	60.00
All Course	42.67	45.08	44.05	41.99	42.33	44.73	43.48

Table 4.7 shows the two classes data set processed by classification algorithm.

Table 4.7

Two Classes Data Set Classification Accuracy Ratio (%)

Course ID	Zero R	Naive Bayes	SVM	J48	DecisionTable	Random Tree	Average by Course
101	95.45	95.45	95.45	95.45	95.45	86.36	93.94
107	82.6	82.6	56.52	91.3	78.26	73.91	77.53
12	95.00	95.00	95.00	95.00	95.00	95.00	95.00
126	94.33	77.35	94.33	92.45	94.33	92.45	90.87
127	96.42	96.42	96.42	96.42	96.42	89.28	95.23
128	88.88	88.88	61.11	88.88	88.88	72.22	81.48
146	100	100	100	100	100	100	100.00
149	71.11	60	71.11	84.44	64.44	77.77	71.48
165	74	66	80	78	80	76	75.67
178	91.66	91.66	91.66	91.66	91.66	87.5	90.97
190	68	40	68	68	64	60	61.33
191	80	76.66	80	83.33	76.66	80	79.44
195	66.66	50	66.66	66.66	66.66	61.11	62.96
204	76.92	61.53	76.92	69.23	76.92	69.23	71.79
206	95.83	91.66	95.83	95.83	95.83	91.66	94.44
28	86.66	46.66	86.66	86.66	86.66	60	75.55
29	79.06	88.37	79.06	100	100	95.35	90.31
33	96.66	96.66	96.66	96.66	96.66	93.33	96.11
34	83.87	83.87	83.87	100	100	90.32	90.32
36	69.23	38.46	69.23	69.23	69.23	61.53	62.82
Average by Algorithm	84.62	76.36	82.22	87.46	85.85	80.65	82.86
All Course	84.21	75.4	84.21	84.21	84.21	83.86	82.68

4.5 Accuracy Ratio Comparison

Table 4.5, Table 4.6 and Table 4.7 show the results of classification processing in prediction accuracy ratio. These results processed by six algorithms. Three data sets class scheme were gathered by course ID and by all course grouping. The data sets prediction accuracy ratio comparison shown in Table 4.8

Table 4.8

Data Set Prediction Comparison by Accuracy Ratio

DataSet Class Scheme	Accuracy All Course (%)	Accuracy by Courses Average (%)
Eight classes data set	23.57	36.48
Three classes data set	43.48	60.00
Two classes data set	82.68	82.86

Table 4.8 compared the accuracy ratio of three dataset schemes. The two classes dataset accuracy ratio is 82.86% that the most highly of all. The accuracy ratio more than 75% is the generally accepted value (Witten & Frake, 2005). Therefore, the two classes' data set is the most suitable for this work data manipulation.

4.6 Learning Activities Grouping

The two classes dataset (Table 4.7) is the suitable data set for this study. This dataset is assembled by 20 activities. The activities are the activities in data mining process for prediction and for activities identification. Hence, the twenty activities appeared in Table 4.3 are the activities that support the high prediction accuracy ratio.

One objective of this study is to identify activities of the log. To understand more relationship detail of these important activities, the general process is factor analysis method. This factor analysis of this data set result shown in Table 4.9.

Table 4.9

Activities Grouping by Factor Analysis Process

Activities ID	Activities Name	Component 1	2	3	4
1	quiz_view	0.934	-0.245	-0.061	-0.049
2	quiz_view_summary	0.931	-0.262	-0.062	-0.026
3	quiz_continue_attempt	0.896	-0.342	-0.067	-0.081
4	quiz_close_attempt	0.879	-0.388	-0.071	-0.039
5	quiz_attempt	0.876	-0.404	-0.073	-0.041
6	quiz_review	0.871	-0.356	-0.068	-0.084
7	url_view	0.718	0.008	-0.02	-0.041
8	course_view	-0.105	0.896	0.014	0.089
9	assignment_view	-0.401	0.855	0.002	-0.059
10	resource_view	-0.124	0.829	0.144	-0.036
11	forum_view_discussion	-0.493	0.763	0.14	0.025
12	assignment_upload	-0.517	0.732	0.009	-0.112
13	forum_add_post	-0.575	0.703	0.125	-0.039
14	forum_view_forum	-0.356	0.695	0.174	0.236
15	forum_delete_discussion	0.005	-0.035	0.967	0.016
16	forum_add_discussion	-0.139	0.075	0.955	0.01
17	forum_delete_post	-0.078	0.212	0.742	-0.069
18	forum_unsubscribe	-0.017	-0.027	-0.01	0.951
19	forum_subscribe	-0.04	-0.059	-0.009	0.941
20	forum_update_post	-0.169	0.435	-0.056	0.518

*Note:
Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
A Rotation Converged in 5 Iterations.

Table 4.9 is the activities grouped by factor analysis process. The result shows four groups of activities that are similar based on statistics. This result was the activities grouped from the same module name and from different module name. Another grouping method, the activities could be grouped by their module name based on the same module type that is shown in Table 4.10

Table 4.10

Activities Grouping by Module Name

Activities ID	Activities Name	Module Name	Group
1	forum_add_post	Forum Activity	1
2	forum_view_forum	Forum Activity	1
3	forum_view_discussion	Forum Activity	1
4	forum_delete_discussion	Forum Activity	1
5	forum_add_discussion	Forum Activity	1
6	forum_delete_post	Forum Activity	1
7	forum_unsubscribe	Forum Activity	1
8	forum_subscribe	Forum Activity	1
9	forum_update_post	Forum Activity	1
10	quiz_view	Quiz Activity	2
11	quiz_view_summary	Quiz Activity	2
12	quiz_continue_attempt	Quiz Activity	2
13	quiz_close_attempt	Quiz Activity	2
14	quiz_attempt	Quiz Activity	2
15	quiz_review	Quiz Activity	2
16	course_view	Assignment & Resource	3
17	assignment_view	Assignment & Resource	3
18	assignment_upload	Assignment & Resource	3
19	resource_view	Assignment & Resource	3
20	url_view	Assignment & Resource	3

Table 4.10 are the activities grouped by eLearning module name. The result shows three groups of activities that are similar by activity modules. These two types of groupings can make predictions based on their activities groupings. The results are shown in Table 4.11 and Table 4.12

Table 4.11

Factor Analysis Group: Classification Accuracy Ratio (%)

Group	ZeroR	Naive Bayes	SVM	J48	Decision Table	Random Tree	Average by Group
Activities Group 1	84.21	84.08	86.78	86.78	86.78	86.78	85.90
Activities Group 2	84.21	82.37	84.21	84.21	84.21	83.52	83.79
Activities Group 3	82.90	82.90	82.90	82.90	82.90	82.90	82.90
Activities Group 4	79.53	75.34	79.53	79.53	79.53	79.53	78.83

*Note:
Activities Group 1: quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review, url_view.
Activities Group 2: course_view, assignment_view, resource_view, forum_view_discussion, assignment_upload, forum_add_post, forum_view_forum.
Activities Group 3: forum_delete_discussion, forum_add_discussion, forum_delete_post.
Activities Group 4: forum_unsubscribe, forum_subscribe, forum_update_post.

Table 4.12

Activities Group (Module Name): Classification Accuracy Ratio (%)

Group	ZeroR	Naive Bayes	SVM	J48	Decision Table	Random Tree	Average by Group
Activities Group 5	83.49	81.35	83.49	83.49	83.49	82.67	83.00
Activities Group 6	86.36	86.36	86.36	86.36	86.36	86.36	86.36
Activities Group 7	84.21	82.83	84.21	84.21	84.21	83.98	83.94

*Note:
Activities Group 5: Forum Activity Group (forum_add_post, forum_view_forum, forum_view_discussion, forum_delete_discussion, forum_add_discussion, forum_delete_post, forum_unsubscribe, forum_subscribe, forum_update_post)
Activities Group 6: Quiz Activity Group (quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review)
Activities Group 7: Assignment & Resource Group (course_view, assignment_view, assignment_upload, resource_view, url_view)

Table 4.11 shows the average accuracy ratio group by factor analysis. All the average accuracy ratios are higher than 75%. Meanwhile, Table 4.12 shows the average accuracy ratio group by module name that all average accuracy ratios are higher than 75% as well. Based on the target rate of predicting performance is determined as at least 75% (Witten & Frake, 2005). Hence, both groups by factor analysis and module name average accuracy ratios are not significantly different.

From the above content, the target of activities grouping is to know the deep relationship of all activities. The results show the relationship of the activities derived from factor analysis and the activities selected from similar names. As a result of this segmentation, predictive learning can reach the hidden relationships between those activities. It can be explained that the finding of relationships between these activities makes predictive learning more effective by comparing the accuracy obtained from all- activities processing and group- activities processing. Comparison of these results is shown in the Table 4.13.

Table 4.13

Comparing the accuracy obtained from all- activities and group-activities processing

Group	ZeroR	Naive Bayes	SVM	J48	Decision Table	Random Tree	Average
Activities Group 1	84.21	84.08	86.78	86.78	86.78	86.78	85.90
Activities Group 2	84.21	82.37	84.21	84.21	84.21	83.52	83.79
Activities Group 3	82.90	82.90	82.90	82.90	82.90	82.90	82.90
Activities Group 4	79.53	75.34	79.53	79.53	79.53	79.53	78.83
Activities Group 5	83.49	81.35	83.49	83.49	83.49	82.67	83.00
Activities Group 6	86.36	86.36	86.36	86.36	86.36	86.36	86.36
Activities Group 7	84.21	82.83	84.21	84.21	84.21	83.98	83.94
All Activities	84.21	75.4	84.21	84.21	84.21	83.86	82.68

Table 4.13 Shown all-activities processing accuracy ratio (by average) is less than almost all group-activities processing accuracy ratio (>82.68). There is only one group-activities processing accuracy ratio (activities group 4) is lower than all-activities processing accuracy ratio. Hence, it could be seen that the method of grouping activities enables the efficiency of predictive learning to be higher.

4.7 Algorithm Comparison

In the past study, the use of data from eLearning systems shows the difference in predictive results with various algorithms. In this study, six algorithms were chosen to test the predictive effect as the guidelines of Ribeiro and Cardoso (2008) , which is shown in Table 4.14.

Table 4.14

Comparing the accuracy by algorithms

Group	ZeroR	Naive Bayes	SVM	J48	Decision Table	Random Tree	Average
Activities Group 1	84.21	84.08	86.78	86.78	86.78	86.78	85.90
Activities Group 2	84.21	82.37	84.21	84.21	84.21	83.52	83.79
Activities Group 3	82.90	82.90	82.90	82.90	82.90	82.90	82.90
Activities Group 4	79.53	75.34	79.53	79.53	79.53	79.53	78.83
Activities Group 5	83.49	81.35	83.49	83.49	83.49	82.67	83.00
Activities Group 6	86.36	86.36	86.36	86.36	86.36	86.36	86.36
Activities Group 7	84.21	82.83	84.21	84.21	84.21	83.98	83.94
All Activities	84.21	75.4	84.21	84.21	84.21	83.86	82.68
Average	83.56	82.18	83.93	83.93	83.93	83.68	83.53

Table 4.14 Shown the comparison of accuracy ratio of six algorithms as ZeroR, Naïve Bayes, SVM, J48, Decision Table and Random Tree. There are three algorithm SVM, J48 and Decision Table show the highest accuracy ration as 83.93 more than Random Tree (83.68), ZeroR (83.56) and Naïve Bayes (82.18). The result is the same approach of other research (Mahboob, Irfan, & Karamat, 2016; Nongkhai & Kaewkiriya, 2015) that the accuracy rations of J48 are greater than Naïve Bayes. Including one more research result that SVM is greater than Naïve Bayes (Ribeiro & Cardoso, 2008). In summary, the results of this study clearly show that the use of SVM algorithm leads to a high level of prediction accuracy which supports the previous research.

4.8 Model Implementation by Two Classed Data Set

The one objective of this work is to construct an eLearning model that could predict the learning outcomes. For this proposed model will be processed on the two classes data set for seven groups of activities (see in Table 4.11 – 4.12). Every time the data set has been processed, an algorithm with high predictive accuracy will be chosen for the process of developing a model. All results of every algorithm that processed by two classed data set can be presented as the tree for programmers to develop the prediction model prototype. There are examples of the prediction results in order to develop a prediction model prototype by tree visualization.

Figure 4.1 shows the learning outcome prediction model for the first- activities group that consists of 7 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of the tree as quiz_continue_attempt, quiz_view, url_view, quiz_close_attempt, quiz_attempt, quiz_review and quiz_view_summary.

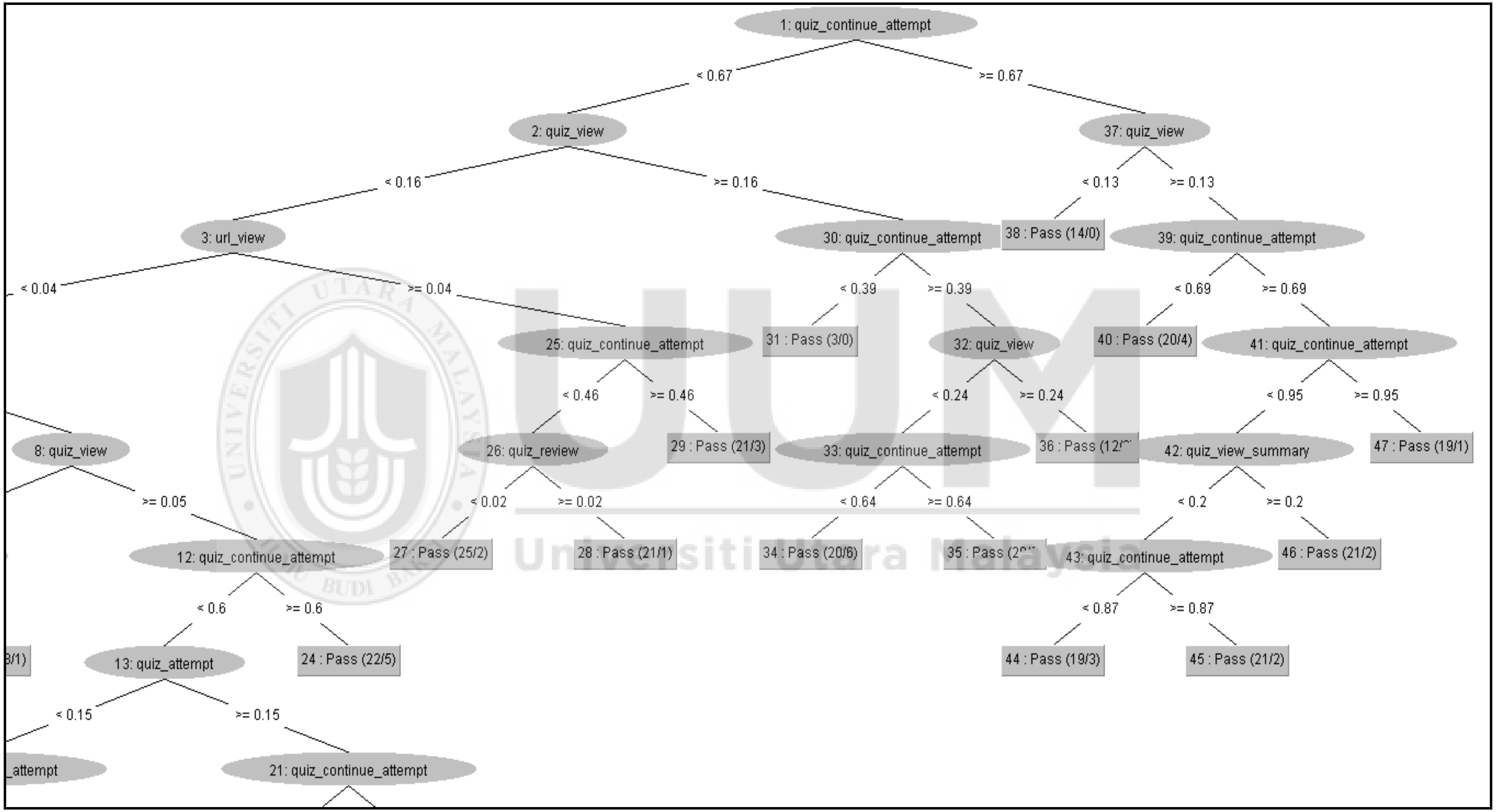

Figure 4.1. Sample of learning outcome prediction model for activities group 1

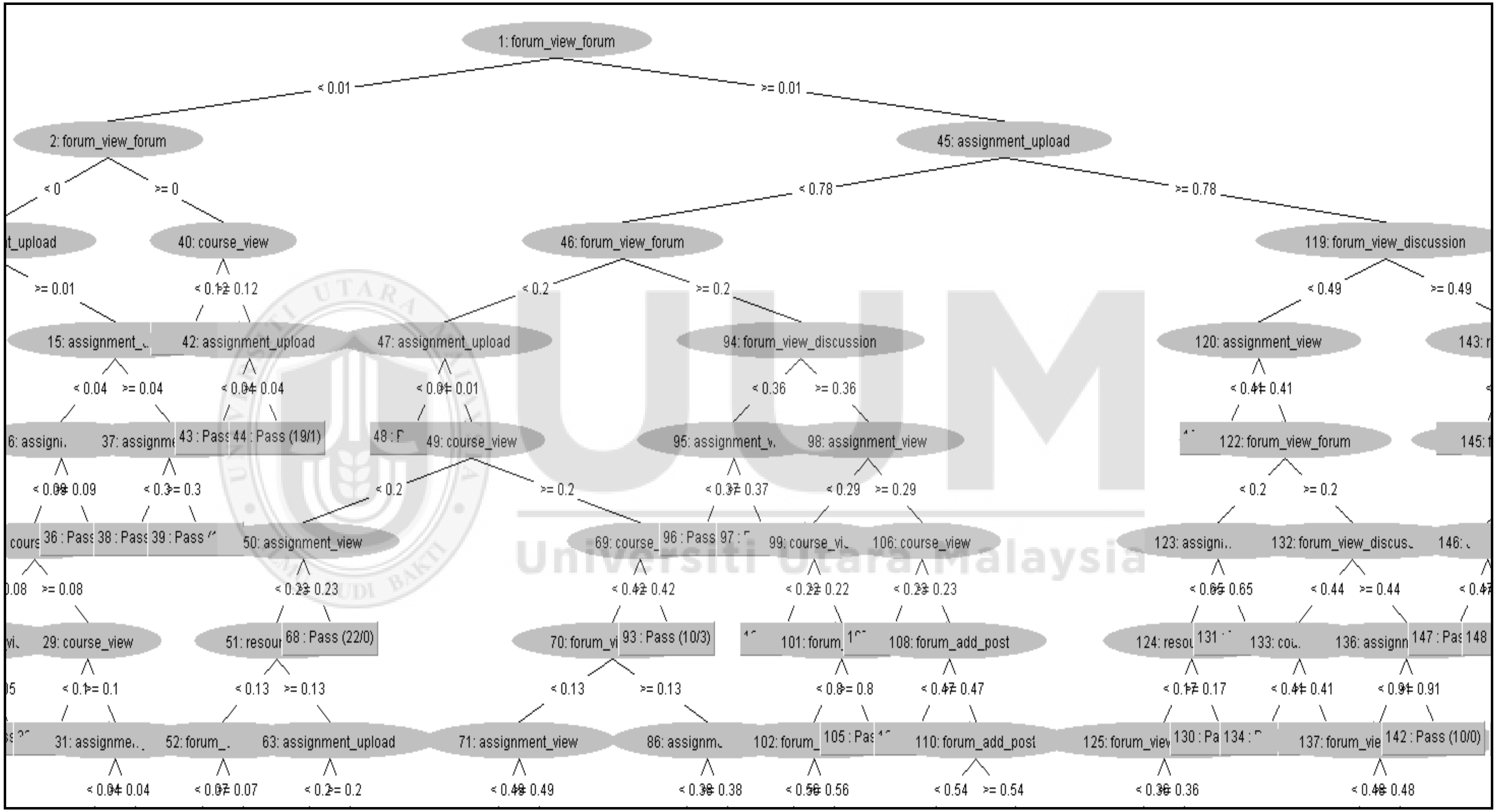

Figure 4.2. Sample of learning outcome prediction model for activities group 2

Figure 4.2 shows the learning outcome prediction model for the second-fact activities or group that consists of 7 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of the tree as forum_view_forum, assignment_upload, resource_view, assignment_view, course_view, forum_view_discussion and forum_add_post.

Figure 4.3 shows the learning outcome prediction model for the third-activities group that consists of 3 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of the tree as forum_delete_post, forum_add_discussion, forum_delete_discussion.

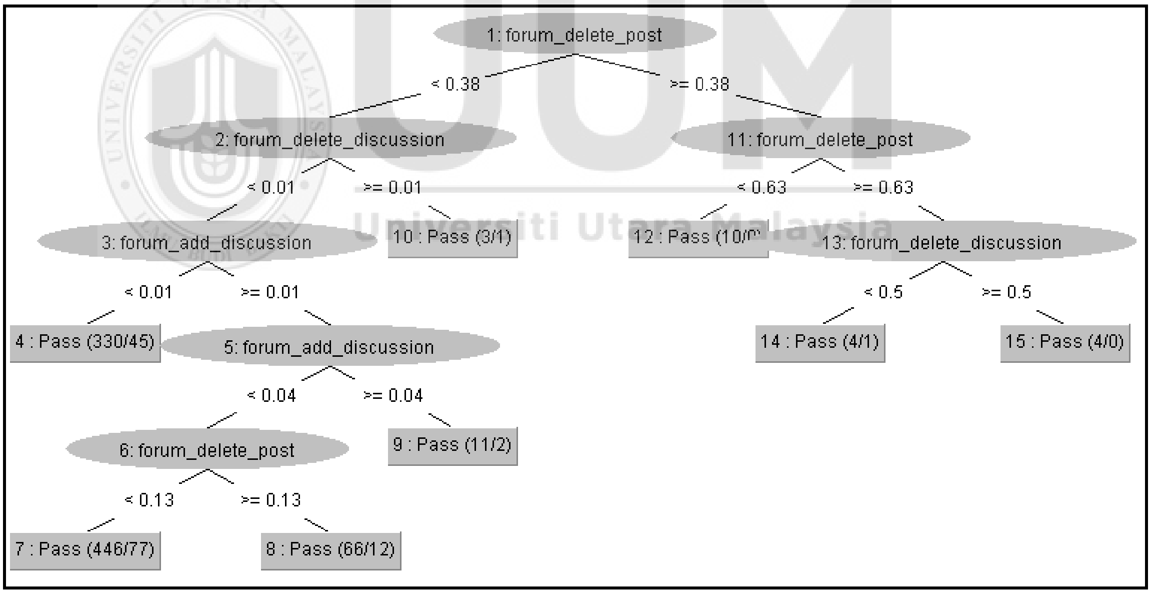

Figure 4.3. Sample of learning outcome prediction model for activities group 3

Figure 4.4 shows the learning outcome prediction model for the fourth-activities group that consists of 3 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of the tree as forum_update_post, forum_subscribe and forum_unsubscribe.

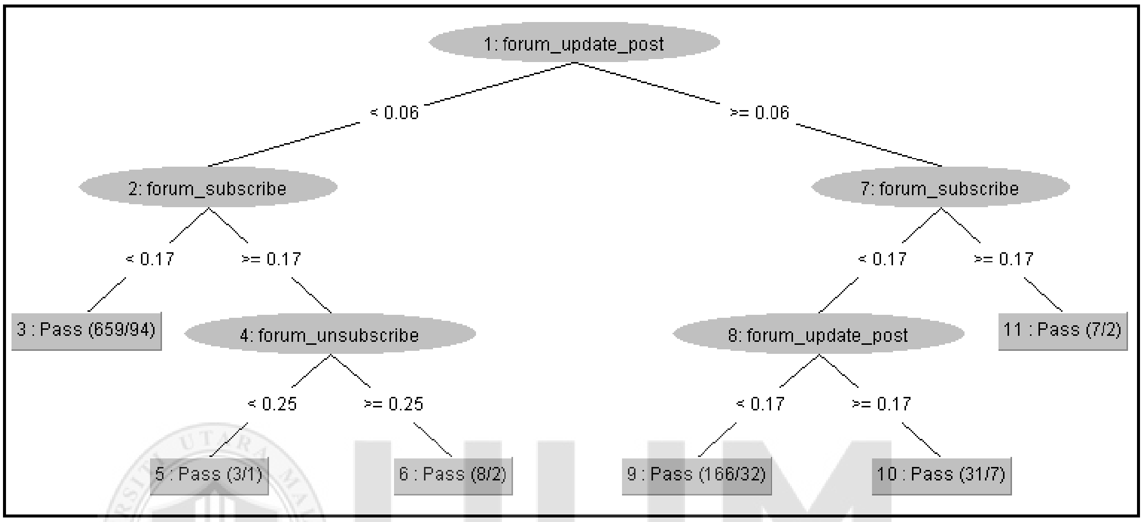

Figure 4.4. Sample of learning outcome prediction model for activities group 4

Figure 4.5 shows the learning outcome prediction model for the fifth-activities group that consists of 9 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of tree as forum_update_post, forum_add_post, forum_delete_post, forum_subscribe, forum_unsubscribe, forum_add_discussion, forum_view_discussion, forum_view_forum and forum_delete_discussion.

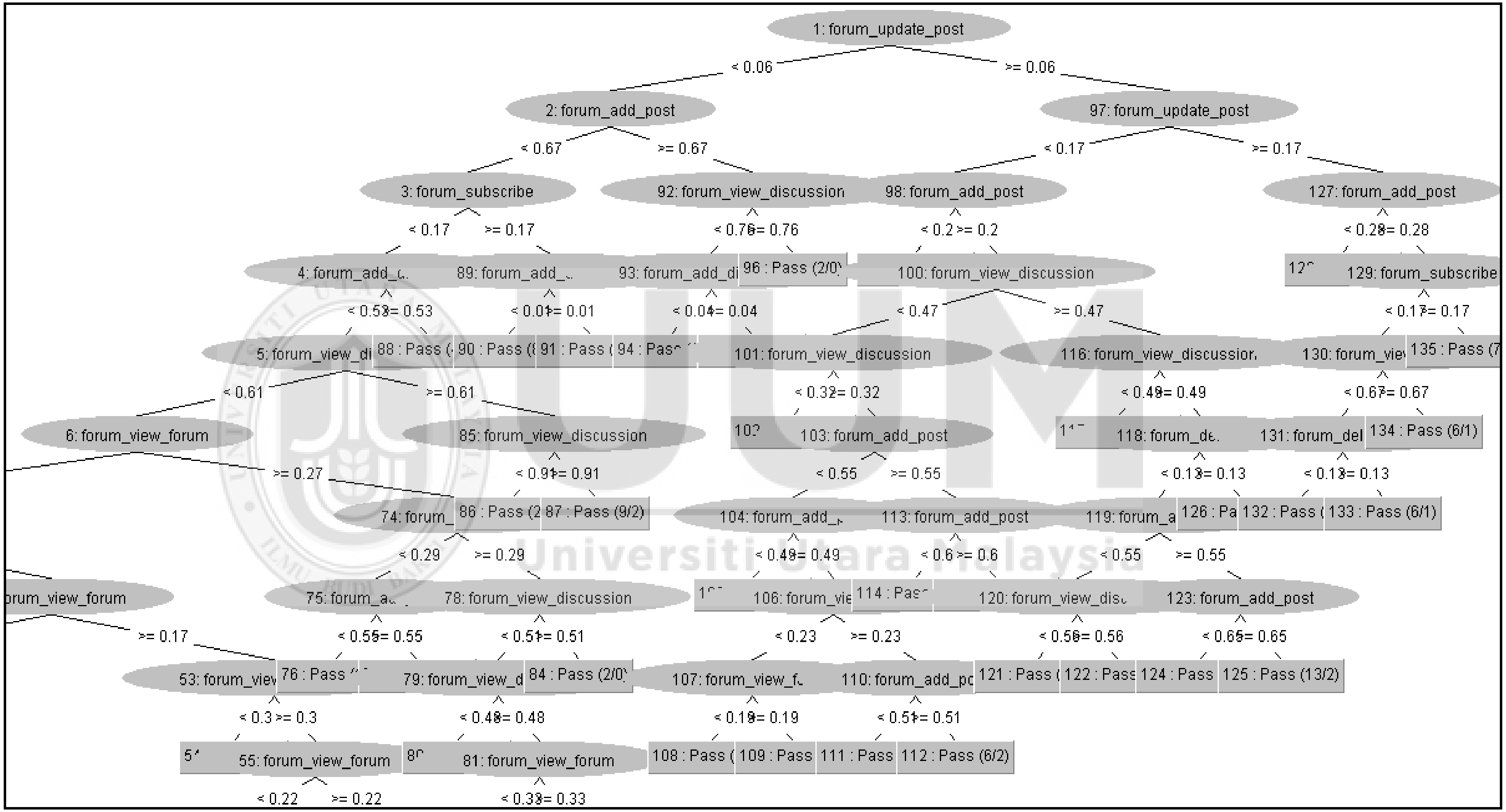

Figure 4.5. Sample of learning outcome prediction model for activities group 5

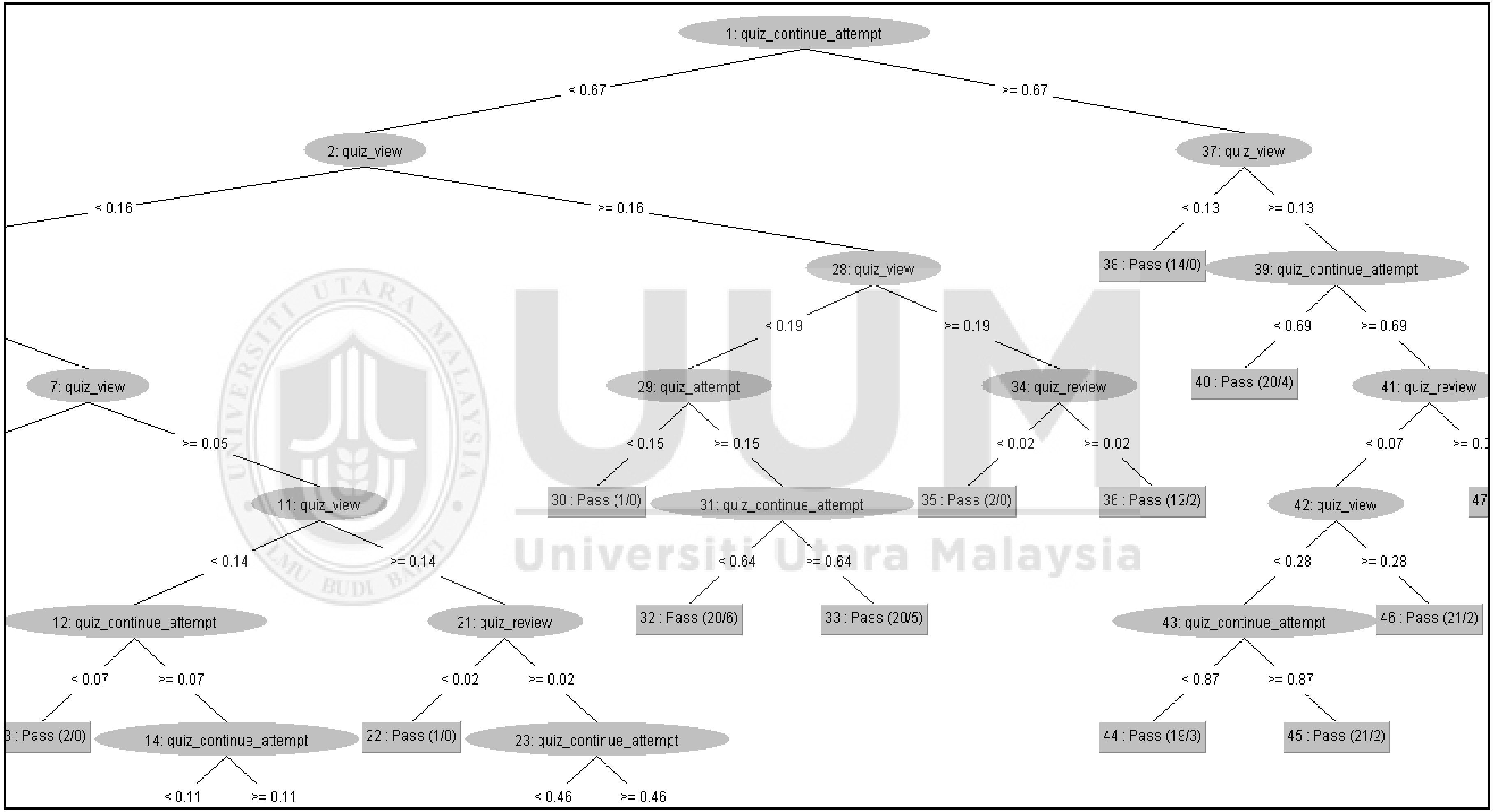

Figure 4.6. Sample of learning outcome prediction model for activities group 6

Figure 4.6 shows the learning outcome prediction model for the sixth-activities group that consists of 6 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of the tree as quiz_continue_attempt, quiz_view, quiz_view_summary, quiz_review, quiz_close_attempt and quiz_attempt.

Figure 4.7 shows the learning outcome prediction model for the seventh-activities group that consists of 5 activities. This model presents the relationship between decision activities and decision number that can order from top to bottom of the tree as assignment_view, resource_view, course_view, assignment_upload, url_view.

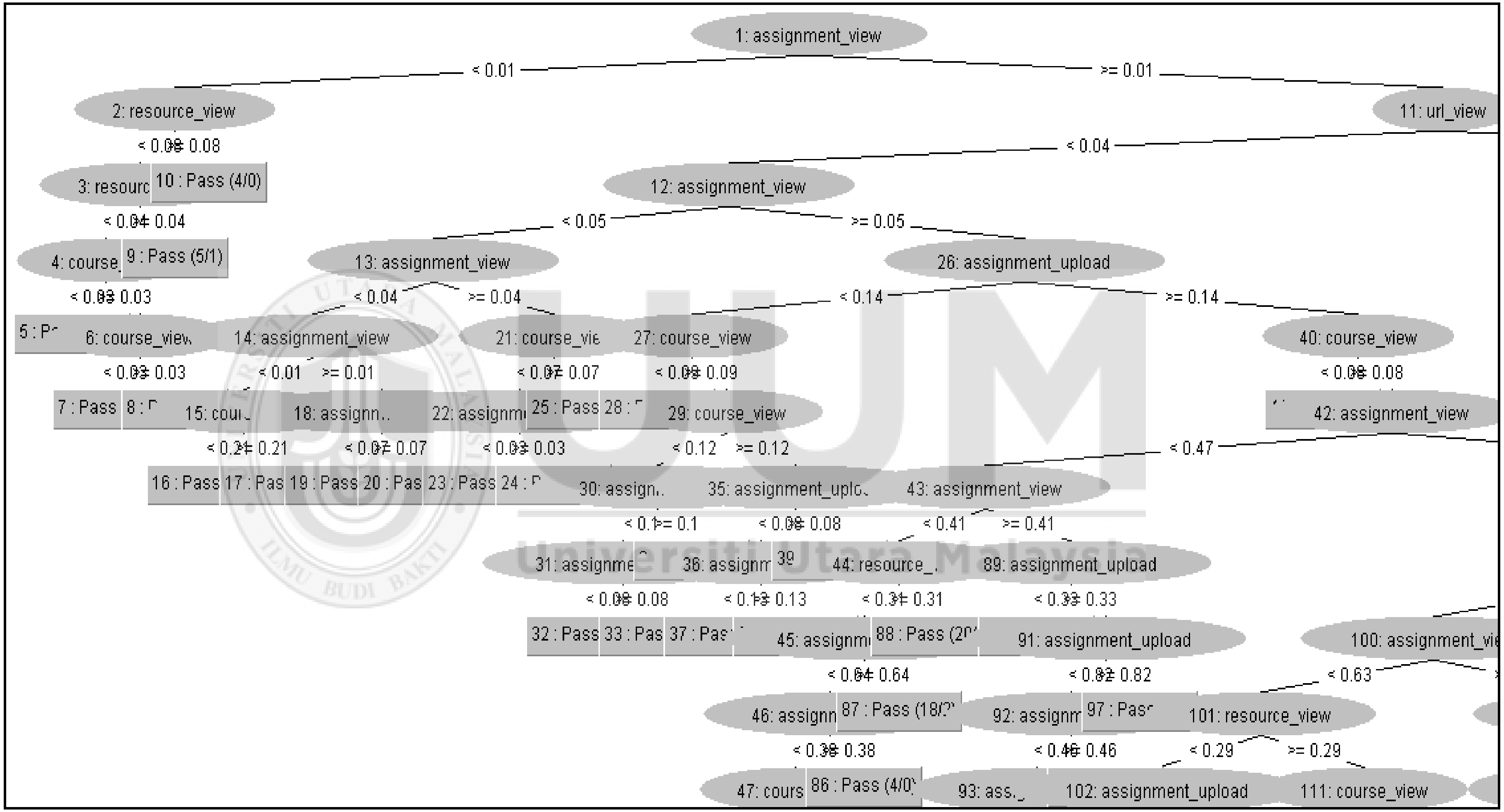

Figure 4.7. Sample of learning outcome prediction model for activities group 7

Figure 4.1 to 4.7 present the models of the relationship of activities in each activities group. The relationship shown in the tree diagram represents the value of each activity, leading to a highly successful learning achievement and low literacy. For these models implementation, we can develop a plugin application for further installation on Moodle eLearning system. Finally, it could be used to predict the learning outcomes of new eLearning activities based on the same eLearning system.

4.9 Summary

Based on this study results, the possibility of a prediction model that can inform the learners and instructors on improving their learning behavior to gain higher learning outcomes is known. The prediction models construction determined by multiple algorithms to explain the accuracy of the predictive effect. Therefore, learning from the log for predicting learning outcomes on an incomplete eLearning system can reach an acceptable learning outcome with an accuracy ratio.

Grouping activities with different methods is a technique that makes predictive results more accurate. In this study, the first objective is to find a method to predict the effects of eLearning on the uncertainty of eLearning activities. This working model is designed to be more neutral in order to create a predictive model for eLearning.

This modelling will respond in the processing part to achieve a predictive model of learning outcomes. At the same time, this work has developed an overview model that will explain how to apply the predictive modelling model to higher education institutions that are using eLearning to invest in education. Therefore, the development of such models will need to evaluate the credibility of the statistics and from the experts in the eLearning implementation as well. The evaluation of these developed models will be presented in the next section.

CHAPTER FIVE
ANALYSIS AND DISCUSSION

5.1 Introduction

The previous chapters presented the research result, including the activities that affect learning outcomes, the relationship of those activities and the appropriate models for predicting learning outcomes that could inform the learners and instructors on how to improve their learning behavior to obtain higher learning results. This chapter demonstrates the reliability of the findings. The model evaluation presents the accuracy ratio, confusion matrix and expert review process from eLearning experts in order to measure the performance of the models.

Based on the data collected, this study finds that the data obtained has a number of activities that are considered appropriate for the use in modeling prediction models. The number of activities selected from the dataset after the cleaning process is 20 activities and the total number of possible activities computed from the Moodle eLearning system is 102. When comparing these activities, the activities used in this study were about 19% of all activities. Based on this information, it is evident that the eLearning systems currently used by both the learner and the instructor have a great deal of potential for doing so.

As mentioned above, the decision to use the activity is based on readiness and limitations as mentioned previously. From this point of view, it is possible that the opportunity to find the appropriate action in the research related to the eLearning system is very possible. However, considering the number of activities that can be used in this research, it is not that great compared to all possible activities. But the results of the analysis can also be expected from the research, which is a group of activities that can be modeled as models for predicting good learning outcomes. It is possible that the estimated data stored in each of these activities is sufficient for the required processing.

Predictive learning in the form of the classification technique is usually appropriate to work with the data set we already know its class number. In other words, it is a data that has a certain amount of results, for example, learning outcomes at the end of a term. Learning outcomes at the end of this term, for example, A, B+, B, C+, C, D+, D, F. The number of these possible outcomes, we call it the class. The number of classes depends on how well each institution determines what the grade is and how well it scores. In the process of machine learning, the quality of processing depends on the characteristics of the information acquired in terms of quality and quantity. Therefore, it is likely that the number of layers of data we want to process in order to predict future results is not the same as the number of layers defined by the institution. Based on the results of the previous study, it can be seen that good predictions of results require a high predictive value. Therefore, the division of learning outcomes may be a natural improvement of the information obtained.

In this experiment, data was passed through the data mining technique to determine the optimal number of classes, two data sets, three dataset classes and eight dataset classes. The result is that the two classes' dataset is the class that receives the highest accuracy ratio. Therefore, from this experiment, we can conclude that the dataset with fewer classes tends to be more accurate in predictive effect.

Another experiment of this research is to find relevant activities and to predict high-level learning outcomes. Comparative testing of learning outcomes from a set of all activities to a set of activities with just a few activities was performed by re-grouping. The newly grouped datasets are divided into two groups: factor analysis and grouping with similar activities names. The results indicate that the categorical activities give the predicted value of the learning outcomes higher than the non-categorical activities. In this experiment, it can be concluded that the acquisition of models predicts learning outcomes, not necessarily from a large number of activities. It depends on the relationship of the activities that play a role in each other. This conclusion supports the fact that learning about good learning prediction on the eLearning system by machine learning is not always necessary to collect the highest number of activities.

5.2 The Research Model

The eLearning instructional components study aims to promote online learning rather than in the normal classroom. This work finds out how to develop this tool to help learners understand the behavior and how to improve the learning activity during their courses. This study could provide the tools that help to predict the learner's learning result before the end of semester. It allows learners to adjust their learning habits to respond to higher learning outcome. The predictive model of learning behavior of learners via eLearning system is a dynamic function. When creating the flexible predictions that respond to ever-changing learning behaviors, eLearning can add value to the learner, the university, and the country education department. It is also concerned with the development of effective teaching and learning as well. This study presents the relevance of the results of the study. This suggests that the development of tools that can enhance the effectiveness of teaching and learning management can have a wide range of effects as the model shown in the Figure 5.1.

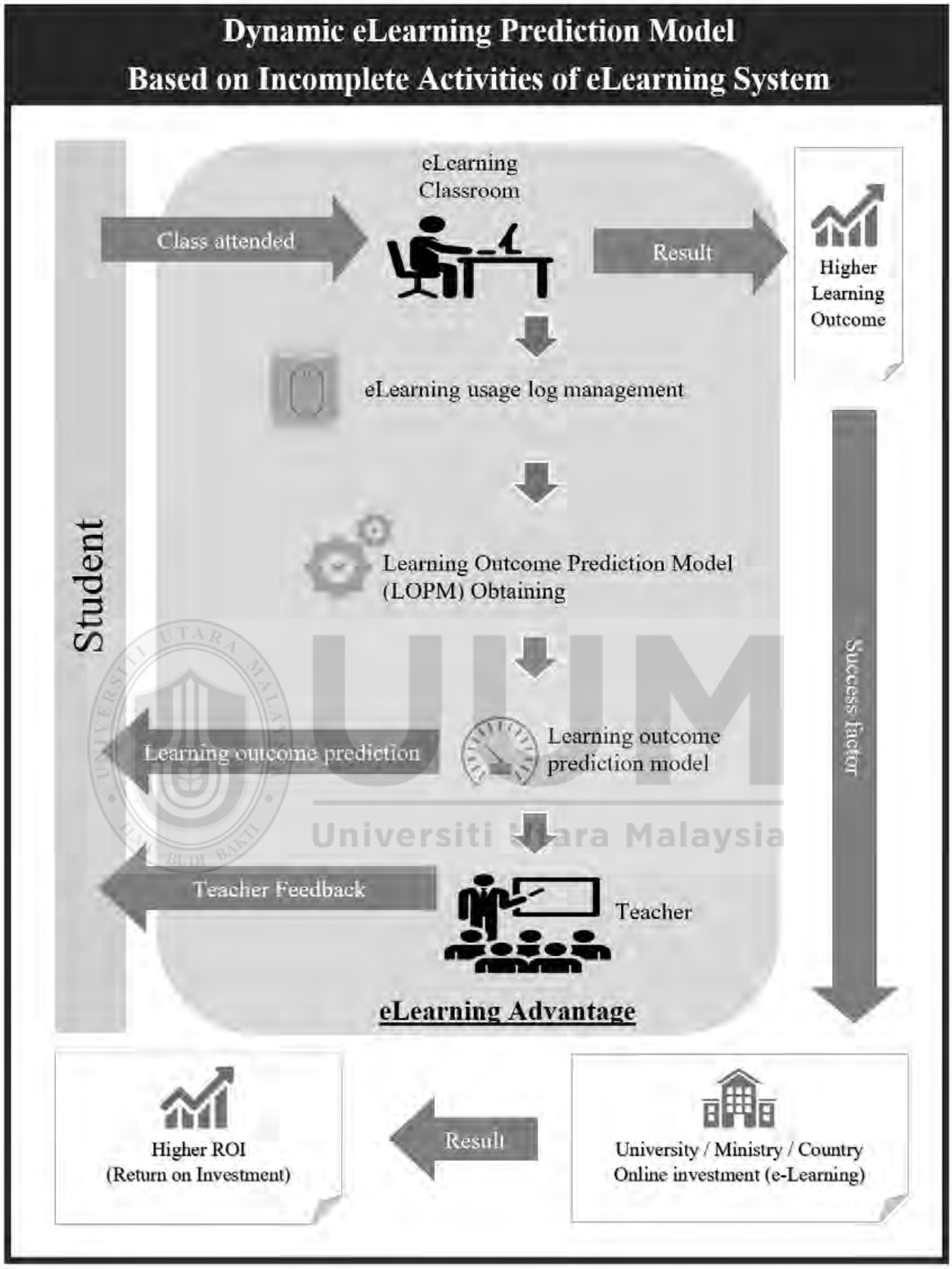

Figure 5.1. Dynamic eLearning prediction model based on incomplete activities of eLearning system

The dynamic eLearning prediction model based on incomplete activities of eLearning system diagram shows the input that the learner accesses the eLearning system and do activities within the system. Then the system records the usage log and processes it to the learning outcome prediction model for learning feedback. The learning outcome prediction model process steps consist of:

a. Data preparation and cleaning ready for data mining process.
b. Activities identification and group of related activities. The reliability of the high-level predictive test should be high.
c. Finding algorithm to make a reliable prediction model.
d. Presentation of learning predictive output to learner and instructor.

These steps of learning outcome prediction model are shown in the Figure 5.2.

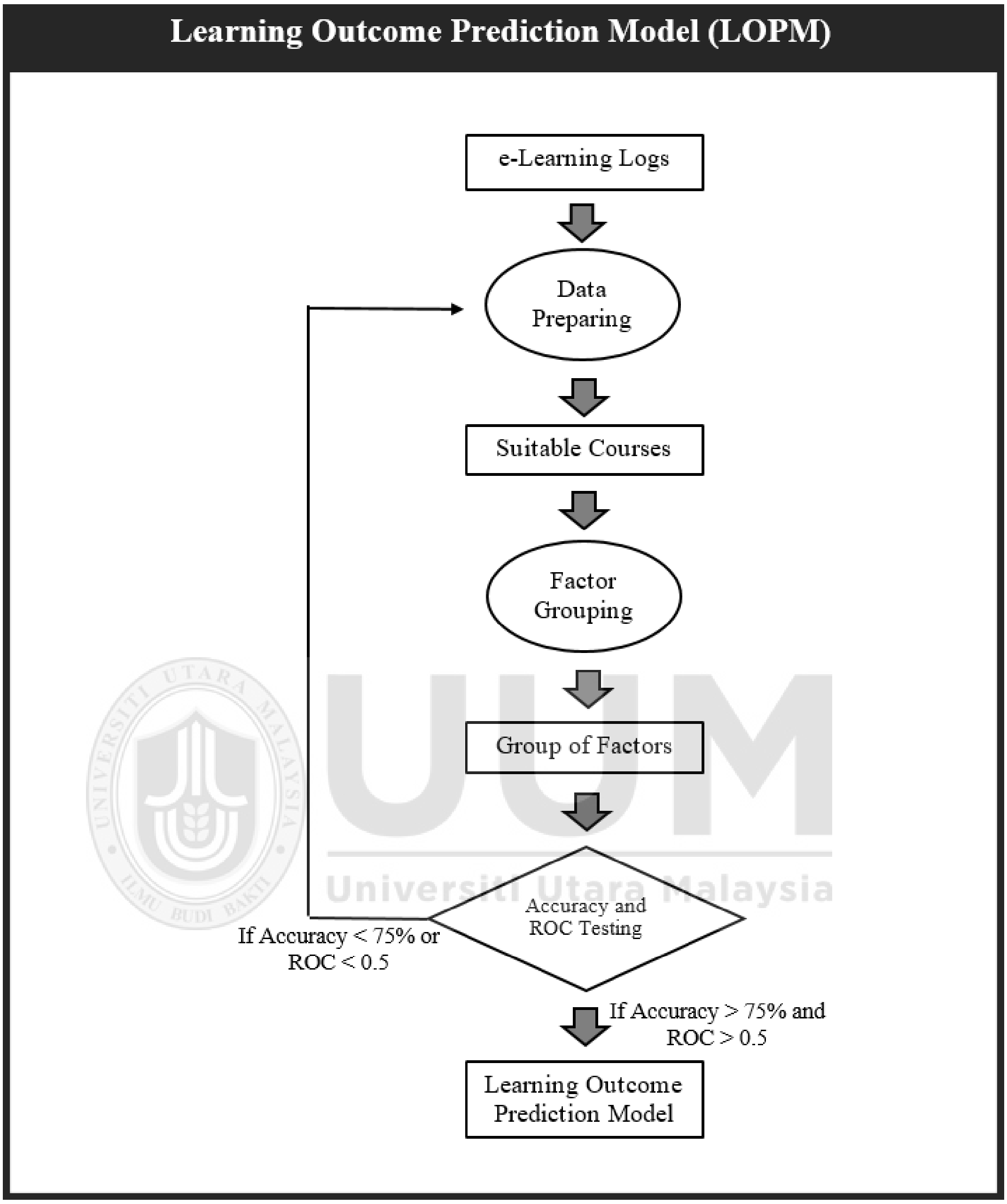

Figure 5.2. Learning Outcome Prediction Model (LOPM)

From the log collection cycle and the predictions of the grades that students will receive at the end of the semester. It may cause students to be eager to change learning behaviors either outside the normal classroom or in the eLearning system.

This behavioral change is an eLearning advantage that is enhanced by regular eLearning. Instructors will be able to streamline the teaching process according to the student behavior and student learning achievement may be obtained at the end of the semester. It can make the learning result higher than normal. This higher level of learning achievement will become a key success factor for supporting the online education investment. That can increase the return on investment (ROI) to universities and the higher authorities responsible for the education of the country.

5.3 Limitation and Delimitation

Data analysis in this study consisted of information on the history of activities, personal data and student grade of each subject. This study uses information from the Moodle learning management system (LMS). We know that the format of the management structure for Moodle is neutral and can be applied to all educational institutions. Hence, the data stored in the Moodle system is a history of student activity, which appears as a log table.

Typically, the log information acquire from the general education institution can be granted by the agency that oversees the use of eLearning. However, the student's personal information and student grade are classified as confidential, even though they are used in the field of research. This may be due to the fact that this information is part of the privacy policy. Normally, all three of these data are not stored in the same university department. Therefore, the acquisition of these two parts requires complex steps to accept. One reason comes from this work researcher who is involved with the department that gives the permission to use these three group of necessary information.

In summary, the need to use all three data are the history of activities, personal data and student grade. The personal data and student grade are the limitation of the data collection in this work.

5.4 Evaluation of Learning Outcome Prediction Model

Predictive modeling works on constructive feedback principle. When we build a model, it gets feedback from metrics, make improvements and continue until we achieve a desired accuracy. Evaluation metrics explain the performance of a model. An important aspects of evaluation metrics is their capability to discriminate among model results. There are several ways to measure the effectiveness of a model. For this work. The result of the prediction is an important part to find its performance. One of the techniques for measuring performance is the confusion matrix, which has the potential to make modeling more reliable and understandable. A confusion matrix is a table that is often used to describe the performance of a classification model (or "classifier") on a set of test data for which the true values are known. For this case, the result of the models was presented the prediction values compared with the real values. Thus, this binary classification outcome activities are suitable to process by confusion matrix method.

From this section onwards, will show the results of model evaluation with confusion matrix method. The activities name of each groups show as below:

Group 1 : Group by factor analysis (quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review, url_view)

Group 2 : Group by factor analysis (course_view, assignment_view, resource_view, forum_view_discussion, assignment_upload, forum_add_post, forum_view_forum)

Group 3 : Group by factor analysis (forum_delete_discussion, forum_add_discussion, forum_delete_post)

Group 4 : Group by factor analysis (forum_unsubscribe, forum_subscribe, forum_update_post)

Group 5 : Group by module name (forum_add_post, forum_view_forum, forum_view_discussion, forum_delete_discussion, forum_add_discussion, forum_delete_post, forum_unsubscribe, forum_subscribe, forum_update_post)

Group 6 : Group by module name (quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review)

Group 7 : Group by module name (course_view, assignment_view, assignment_upload, resource_view, url_view)

Table 5.1

Seven Activities Groups Confusion Matrix

Activities Group	TP	TN	FP	FN	F-Measure (Pass)	F-Measure (Fail)	Accuracy	ROC
Group 1	0.997	0.000	0.842	1.000	0.913	0.000	0.839	0.510
Group 2	0.997	0.000	0.842	1.000	0.913	0.000	0.839	0.440
Group 3	1.000	0.000	0.842	0.000	0.914	0.000	0.842	0.476
Group 4	1.000	0.000	0.842	0.000	0.914	0.000	0.842	0.493
Group 5	0.997	0.000	0.842	1.000	0.913	0.000	0.839	0.427
Group 6	0.999	0.000	0.842	1.000	0.914	0.000	0.841	0.504
Group 7	0.997	0.000	0.842	1.000	0.913	0.000	0.839	0.440

Table 5.1 shows the evaluation of confusion matrix on 7 output models. The evaluation results are considered optional by accuracy ratio and ROC. Table 5.2 compares these options for the model performance ranking.

Table 5.2
Models Performance Ranking

Model name /Confusion matrix option	**ROC Area**	**Accuracy Ratio**	**Ranking**
Activities group 1	0.510	0.839	1
Activities group 2	0.440	0.839	5
Activities group 3	0.476	0.842	4
Activities group 4	0.493	0.842	3
Activities group 5	0.427	0.839	6
Activities group 6	0.504	0.841	2
Activities group 7	0.440	0.839	5

Table 5.2 shows the models performance ranking of 7 output models. When considering ROC Area results, generally values that can be considered should be more than 0.5. Therefore, there are two models' activities group 1 and activities group 6th get ROC Area results more than 0.5 that present their status are low discrimination level. Meanwhile, the accuracy ratio results of all models are higher than 0.75. Hence, all models are the standard for prediction ratio.

One observation of comparing the ranking performance with the accuracy ratio and the ROC is that the accuracy of the ration is high, above 80%, indicating that each model has accuracy in predicting results is high.

Based on ROC values, there are only two activities groups that are greater than 0.5, but less than 0.6, which is considered to be low. The result is that this model may not be reliable enough to predict the future results of other sets of data in the eLearning system. The reason for this is that ROC values are indicative of the reliability of predicting the results in other datasets.

Overall, this study found the performance models that could enable the learning outcome prediction model for supporting the dynamic eLearning prediction model based on incomplete activities of eLearning system which is the ultimate model (Figure 5.1) of this study.

5.5 Expert Review

The model performance ranking as shown in Table 5.2 is the confirmation data that could support this research model "Dynamic eLearning Prediction Model Based on Incomplete activities of eLearning System" as shown in Figure 5.1. However, this research model assembled with various activities that related to learners and their learning result. The study from the expert who concerned in eLearning system could gain more perspective for this research model. Hence, this study determined the expert review process for the research model evaluation as well. For this case, the experts are the people who are working on eLearning assessment and quantitative data analysis for more than ten years. These experts such as the eLearning system administrator or lecturer who used the eLearning system.

The expert review question document is shown this research model (Figure 5.1, 5.2) and it detail to eight experts with the research model (Figure 5.1, 5.2) and the details, including four main questions as below:

1. The components of the model are complete, which can be explained the modeling, predicting learning outcomes and the advantage. (Question no.1)
2. "Dynamic eLearning Prediction Model Based on Incomplete activities of eLearning System" is accurate in the process of developing the learning outcome prediction model. (Question no.2)
3. This "Learning Outcome Prediction Model" effective enough to predict the learning outcomes? (Question no.3)
4. What should be added to the "Dynamic eLearning Prediction Model Based on Incomplete activities of eLearning System" (e.g. model development, performance measurement, model impact study)? (Question no.4)

The expert review answer document designed in two parts. The first part is the expert agreement level for the question number 1-3 by Likert Scale. The second part is the expert discussion by open-ended answer for the question number 1-4.

In each part of this interview, it allows this research to visualize the model that needs to be presented in terms of both scale-level and open-ended comments to further understand other activities.

5.5.1 Expert Agreement Level

Table 5.3

Expert Agreement Level for the Question no. 1-3

Expert no.	Agreement Level (1-5)		
	Advantage	Accuracy	Effectiveness
Expert 1	4	4	4
Expert 2	5	5	5
Expert 3	5	5	5
Expert 4	3	4	3
Expert 5	4	5	5
Expert 6	3	4	3
Expert 7	5	5	5
Expert 8	4	5	5
Average	4.13	4.63	4.38

Agreement level label: 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree

Table 5.3 shows the agreement level of advantage is 4.13 (agree), accuracy is 4.63 (agree) and effectiveness is 4.38 (agree). It is concluded that the experts agree with all three questions by average.

5.5.2 Expert Discussion

Question no.1 “The components of the model are complete, which can be explained the modeling, predicting learning outcomes and the advantage” was discussed by eight experts that model come from the real log that could be the prototype for learning effective comparison and could explain support for learning advantage policy. Meanwhile, there are some questions about other activities that should be related with this model and make it more credible. The ROI should be explained for more specific scope of return of investment detail.

Question no.2 “Dynamic eLearning Prediction Model Based on Incomplete Activities of eLearning System” is accurate in the process of developing the learning outcome prediction model” was discussed by five experts that the accurate result of this model is standard and could predict the learning result. The evaluation process, using the area under the ROC curve is a good evaluation. The average of 2-3 times validation performing can be used to check accuracy is the additional suggestion from expert as well.

Question no.3 “Is Learning Outcome Prediction Model effective enough to predict the learning outcomes?” was discussed by three experts that the model is effective enough and the result should be compared with more algorithms.

Question no.4 “What should be added to the dynamic eLearning prediction model based on incomplete activities of eLearning system? (e.g. model development, performance measurement, model impact study)” was discussed by six experts that future development should be working on smartphone as well. This model focus on student individual learning in the future research should focus more on group learning. This study should apply this model to the easy implementation usage. There are many evaluation techniques that could be considered in the next study. The learning outcome prediction results should go to supervisor, lecturer or counselor in order to give advises to students. The result model should explain more detail for the benefit of e-Leaning implementation. The future study should acquire the other effect from the eLearning system.

5.6 Summary

The reliability of the findings presented the models' performance ranking by ROC area and accuracy ratio (Table 5.2). The activities group 1 (quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review, url_view) could be the highest performance significant model group. At the same time, the expert reviewing for this research model, "Dynamic eLearning Prediction Model Based on Incomplete Activities of eLearning System", mostly agree with it and supported the model idea. However, some of the experts suggested that many evaluation techniques could be considered in the next study, and it should acquire the other effect from the eLearning system as well.

As mentioned in the previous chapters, most of the eLearning systems result in the incomplete activities of eLearning usage, which are caused by differences in the selection of existing tools. The model presented in this research "Dynamic eLearning Prediction Model Based on Incomplete Activities of eLearning System" will provide a solution for selecting those incomplete activities to create a predictive model that is still highly effective. Learning Outcome Prediction Model (LOPM) does not specify based on some of the incomplete activities, but it can find the appropriate incomplete activities for better prediction model construction. In other words, the results of this study enable us to control the selection of incomplete activities or their groups that can be used to create highly effective predictive models and will reduce the limitations of incomplete activities of eLearning usage.

CHAPTER SIX
CONCLUSION AND FUTURE WORK

6.1 Introduction

eLearning is very important in the management of educational institutions. As far as computer technology is concerned, the potential is significantly higher, while its costs are much lower than in the past. For software technology, it has a more artificial intelligent (AI) direction, making the ability to manage the content of the old software more intelligent, as well as the eLearning.

Previously, the development of eLearning was to replace the teaching and learning management of some courses or in many subjects of the university. But the goal of eLearning creating nowadays is not just substituting classroom teaching, it can also cover the learning needs of many people over campus, based on the concept of Massive Open Online Courseware (MOOC).

There are several studies that attempt to study eLearning's work in order to maximize its ability to fully benefit learners. Since eLearning is a learning system, so there is a core of information technology, and the other two essential parts are the learner's and the instructor's part. In order to understand the eLearning, it must be understood in every part of it.

The emergence of data in the eLearning system is very much from all parts that mentioned above. This information is like recording everything that happens in the whole learning process, such as instructional media, activities, and instrumentation. At the same time, the behavior of accessing all things in the learner's system is kept. Hence, in order to learn the stories within the eLearning system for the development of its effectiveness, it is necessary to rely on its vast information.

In view of software development, the eLearning system is a tool that was created to handle all the traditional classroom replacements, while several educational institutions have commented that eLearning is part of the educational investment. Generally, we called this learning management tool as LMS. The currently LMS available are abundant for the optimum use of educational institutions. Most of the LMS discussed here are the open source LMS, especially Moodle, which is one of the most commonly used. There are also many research studies based on Moodle, as well as this research. In terms of impact on the use of eLearning, it will be a great inspiration for spending personal time learning and working with both the learner and the instructor. In addition to the added value of managing time spent on LMS, some systems have also developed tools for creating online communities for sharing and further research from those communities.

In some countries it is found that eLearning reduces the economic gap. Other effects such as reducing the cost of teaching and learning because they do not want to create a real classroom anymore. Another advantage of it is that it is accessible anywhere and anytime, which increases the likelihood of meeting between the learner and the instructor. In higher education, eLearning has also become a part of the education strategy. The current levels of using eLearning range from 1 percent up to 60 percent, and future predictions are from 5 percent up to 75 percent (Wardaya & Pradipto, 2017).

The aim of eLearning is to improve the efficiency of learning tools that help users get more experience and benefits. In the past, we have studied various activities of learning activities on eLearning systems, most of which are open source systems. Learning activities on these open source systems will allow faster and easier cognitive enhancement because of the standard of similar information formats.

The same eLearning systems have the same management tools, such as user management tools, media management tools, communication management tools, test and measurement tools, and so on. These tools will store data while being used in different ways. All these data will be used to learn behavior that will affect the effectiveness of the learning process.

Based on the previous studies of the optimization of the eLearning system, it was found that the required data collection was planned before activating the system. This research method will allow the researcher to control the desired activities and to access the analysis of the relevant data according to the research problem, due to the completeness and sufficiency of the data.

Unfortunately, the reality of the eLearning system is very limited and problematic. Previous studies have investigated the causes of incomplete eLearning applications, such as the attitudes of learners and teachers to eLearning systems that cannot be compared to traditional learning. The fear of using the technology, lack of skills to prepare the appropriate course of the instructor and good administration of the system controllers are all important causes for learning management problems through eLearning. As a result of this problem, the data generated by the eLearning system is very variable. Due to complex user behavior and the large amount of data stored in the system, learning the hidden stuff of these data is more difficult. Therefore, the use of machine learning techniques to manipulate these data is interesting because of its high capacity and flexibility to analyze data. eLearning data mining represents a field of research relevant to the application of data mining, high-level machine learning and statistics-based data generated from the learning management system environment. Knowledge is hidden in the learning environment and can easily be separated through data mining and predictive methods.

One of the most important issues in studying the models of student learning prediction through eLearning data analysis is the lack of flexibility in the implementation. These prediction models are made by defining the scope of the activities that were originally designed, which would allow the collecting process to be simple and sufficient. Then, this process makes it impossible to extend the effect of predictions to cover a large number of activities contained in an eLearning system.

As mentioned above, the completeness of the data required to study the model predicts the learning outcome depends on many activities that contribute to the completion or not. In the eLearning system, learners can do a variety of activities according to each design, depending on the instructor's lesson plan. In each course there is a lot of student activity, which later becomes the stored data of each activities.

On the other hand, the number of activities that occur in each course varies with the activity of the instructor. So, in the course of a student's attendance, it is impossible to predict how many people will use the instrument, at what frequency. This is a complexity that cannot be controlled in various subjects.

In summary, the actual state of the eLearning system is very different in terms of activities selected, different instructors and different learner behaviors. These uncertainties are a real manifestation of the eLearning system used in all educational institutions. Therefore, in order to learn its information to come up with a model for predicting learning outcomes, the data should be used in real state rather than using data derived from specific control activities.

This research explores the ways in which the learner's predictive model can be used to support a variety of activities, and at the same time, these models can be used to predict learning outcomes for other subjects. Therefore, machine learning is the suitable choice for managing such data on this work.

This study defines the concept of research as the study of the eLearning activities affecting to high learning outcome predictable based on incomplete activities of eLearning system. This study will not be specific to any course and will not impose any activities that affect learning. Based on this approach, we have the opportunity to find new activities that affect learning and providing opportunities to model predictive learning outcomes for predicting the future courses. At the same time, this study also aims to provide an overall model that can be applied to the development of eLearning systems in other educational settings as well.

For this study, the meaning of incomplete activities of eLearning system could be explained as the activities designed by courses instructor that become the number of eLearning activities. In fact, the number of activities that students take may be less than the total number of activities available in the course. On the other hand, some of the activities that are available to the user, but too few data of them can be used in the machine learning process, will be meaningful as well.

The research study has four questions: the first question is "what are the important activities that support high outcome prediction?" This question is the first goal of this study to determine whether the history data stored on the incomplete usage eLearning system is ready for use in predicting the outcome. Another is the need to know how deep the relationship between activities is and how it will affect the higher predictions.

The second question is "how can an eLearning outcome prediction model be constructed based on the analyzed eLearning activities?" From the previous questions, we get a group of activities that affect learning outcomes. These activities will be used as a selection process for acceptable activities that can be used to predict learning outcomes. Finally, the process becomes a model for use in screening the appropriate activities for use in predicting the outcome.

The third question is “can an eLearning usage model be synthesized based on incomplete activities?” A study to create a model-finding process that can provide a good predictor of learning outcomes is another important goal. Not only did the process find the activities, but the use of predictive effects to respond to changing learning behaviors was also part of the process. Finally, if these processes can be defined from beginning to end, then the organization will succeed in using the eLearning tools for educational management. Profit success is another opportunity that is in the interest of many higher educations.

The fourth question is “is the model produces acceptable agreement level on advantage, accuracy, and effectiveness?” The overview model found in this research was designed to help the incomplete eLearning system to find the activities used for modeling predictive learning. Including the value of predicting learning outcomes to the learner and the instructor as well as the institution. Is the whole model appropriate for the development of the potential of eLearning applications for other systems? This is a must for experienced professionals to guide and evaluate.

Based on these research questions, the research has been designed to define the scope of the research to be appropriate to the time available, as well as the information that is needed. Then, the research scope of data collection period was six semesters during 2012-05-25 to 2015-04-06. The eLearning web log data set and user profiles were the necessary data from one Moodle eLearning system.

From the data collection plan for this study, the data contained courses, user profiles and log files. These data have been processed through these steps.

Data preprocessing step, the data set obtained from the data collection step has been processed with four stages of preparation as data fusion, data cleaning, data structuration and data summarization. The target of this process is to streamline the data and be ready to be introduced into the data mining process.

eLearning activities analyze step, this step brings all the activities derived from the previous step to processing with the algorithm in the group of classification are ZeroR, Naive Bayes, SVM, J48, DecisionTable and RandomTree. Why choose a classification technique, because the nature of the information used in this study is clearly targeted at the level of academic achievement. Classes considered in this study were conducted in three types as eight classes, three classes and two classes. The result of this process is that some activities have high predictive value by accuracy ratio. The use of two classes gives the highest predictive value by accuracy ratio.

Learning outcome prediction model construction step, on the need to learn the in-depth relationship of activities derived from previous steps, activities group management techniques are a guideline for finding such answers. In this phase, the statistical capabilities of grouping by factor analysis and the process of grouping with similar activities names are used. These two methods are part of the presentation of the process of grouping related activities. As a whole, this step is to find the activities that affect the most predictive learning outcomes in this hierarchy so that they can go back to prepare different information. In summary, this step involves selecting the activities that have been selected and then dividing them into groups and processing them with different algorithms. The predicted outcome if it reaches an acceptable level using the accuracy ratio and ROC is the basis for determining whether it is appropriate to use it. If a given value is reached, it can be used, but if it does not meet the criteria, return to the procedure of preparing the data and re-group the activities so that the appropriate value can be found. This is shown in Figure 5.2.

Dynamic eLearning prediction model based on incomplete activities of eLearning system synthesizing step, this model aims at how to bring the ability of learning outcome prediction model from the previous step to link with the learning eLearning system to maximize efficiency. This is a compilation of many components to synthesize into a snapshot model of all those components. Linking eLearning relevance to the learner and the teacher, the emergence of data to the acquisition of activities and activities affecting learning outcomes. Including the learning outcome prediction model from the previous step is another element under this model. Reviewing other activities related to the implementation of the eLearning system in higher education aims to achieve a higher efficiency of the eLearning system that will respond to the learning outcomes of the learner. Creating a learning-enhancement tool for the eLearning system will provide feedback that will allow learners and instructors to adapt to the learning activity that leads to higher learning success.

From the synthesized process, the result is a snapshot model named "Dynamic eLearning Prediction Model Based on Incomplete activities of eLearning System". Its advantage is not only to predict the learning outcomes of the learners in the same system, but also to become the prototype of this research for use in other eLearning systems as well.

Model Evaluation step, based on the output model "Dynamic eLearning Prediction Model Based on Incomplete activities of eLearning System" this step is for improving eLearning system to add value to both the learner and the instructor as well as to become the prototype for other eLearning systems. It is necessary to evaluate this model in various dimensions in order to confirm its expectations. This evaluation is evaluated by experts who are experienced in the use of eLearning systems in various aspects. The evaluation is divided into assessing the effectiveness of the learning outcome prediction model predictive model using the confusion matrix process to determine its reliability. As part of the overview model, the use of expert opinion was divided into Likert scale method and open-ended questions. The main question of this assessment is to confirm experts' opinion on the issue of advantage, accuracy and effectiveness as well as other expert comments to be offered in the form of open-ended questions.

Based on the research methodology above, each of these steps has generated results from the processing of the research questions. These results be able to address the objectives of the research, as will be shown in the next section.

6.2 Achieved Objective

The objectives of this study are the four main issues to process. The first objective is to analyze the activity that affected high learning outcome prediction. For this study data set, there are 53 courses provided by eLearning. After log cleaning processing and matching to the available grade result there are 20 courses and 20 activities remaining that can be processed. The classification process was run by several algorithms based on eight classes, three classes and two classes data set. This dataset prediction accuracy ratio of two classes is 82.68%. Then, the two classes data set is accepted by more than 75% accuracy ratio. The process was continued by factor analysis and similar name grouping. The result found 4 significant groups by factor analysis process and 3 groups by similar name grouping process. The prediction accuracy ratio for each group are more than 75% by average. The evaluation model by confusion matrix method was found two group that the ROC (Receiver Operating Characteristics) value more than 0.5 (considerabel value > 0.5). Therefore, the activities group (quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review, url_view) was the activities that could be the significance for learning outcome prediction model construction. However, these activities are only the result of processing for this dataset. If there are additional system using information in the future, these activities names may change. As explained above, the concept of this research does not find specific activities for use as a permanent activities, but rather as a temporary activities that can explain acceptable values of predictive learning. Therefore, it may be possible to find the good models for predicting learning outcomes and to present the relationships of different activities sets. An example of a set of activities that can predict good learning outcomes is shown in Figure 4.1. In conclusion, the first objective of this study was achieved as mentioned.

The second objective is to construct a learning outcome prediction model (LOPM) for eLearning usage. The process of the model gaining is data preparing to find the suitable courses for doing the prediction task by data mining technique. The activities grouping to find more accuracy ratio is the next important process. The last process is activities group checking for the reliability. Then the process will get the good activities that suitable for the future prediction model processing. For this study, the accuracy ratio higher than 75% and ROC (Receiver Operating Characteristics) gets value between 0.5 and 0.6. All model completed by accuracy ratio but only two models pass 0.5 at the low discrimination level. Hence, this could build the learning outcome prediction model for eLearning usage. Overall, the second objective of this study was achieved.

The main step is to find the activities and segment them based on the acceptable accuracy ratio and low discrimination value of ROC. On the basis of this data set, looping to find the appropriate set of activities is a group of seven models. Finally, there are only two models that have both the correct ratio to be used as a model for predicting good learning outcomes. For the results, learning outcome prediction model LOPM shows that the concept of finding a activities by this selection process is possible.

The third objective is to synthesize a dynamic eLearning prediction model based on incomplete activities of eLearning system. This objective aims to propose to relationship of the eLearning determinant whether learning outcome prediction model is acceptable or not. The determinant of dynamic eLearning prediction model based on incomplete activities of eLearning system are the student, teacher, eLearning logs, learning outcome prediction model and the eLearning advantage. In addition, the modeling of these elements is based on previous research review.

As have been discussed, as well as the review of previous research, the model presents the interconnectedness of each other in a systematic way. Therefore, this model is a systematic model that introduces the early point of introducing the eLearning system as a tool for learning management of educational institutions. The model then uses learning LOPM as a tool for predicting learning outcomes and forging predictive results to learners and instructors. The model suggests that if the system can help learners and instructors know the learning outcomes in advance, this prediction tool will change the behavior before the end of the semester. If this looping affects effective learning, the next step is the success of the institution's investment in bringing the eLearning system to use as an educational aid. Finally, this work can create models that illustrate the introduction of machine learning to add value to the eLearning system as described.

The fourth objective is to evaluate the dynamic eLearning prediction model based on incomplete activities of eLearning system on advantage, accuracy and effectiveness. The model determinants were considered by experts' opinion. The results of experts reviewing were agreed with all three questions by agreement level of advantage is 4.14 (agree), accuracy is 4.57 (agree) and effectiveness is 4.29 (agree). It is concluded that the experts agree with all three questions by average (4.33).

As mentioned above, the dynamic eLearning prediction model based on incomplete activities of eLearning system is a system model that offers the ability of machine learning to be used as a tool to help existing eLearning systems have higher potential. This higher level is due to the fact that learning outcome prediction model (LOPM) found a set of activities that can be modeled as predictors of learning outcomes. From both the learning outcome prediction model (LOPM) and the processes used in the whole model, the expert has the opinion that it is the highest accuracy, followed by the effectiveness and advantage. Overall, the opinions in the various aspects are at the level of agree (4.33 by average) that the value of interest. At the same time, most of the comments from the open-ended questions section are encouraging the details of this model.

6.3 Implication

The main purpose of this study is to increase the potential of the eLearning system. Basically, eLearning has the ability to substitute the normal classroom and increase the learning time for the learner. Currently, the new information technology owned by most people are laptops or smartphones. These devices became part of life. It also makes eLearning more accessible to the general public and not specific to the students anymore.

According to this study, the results found the learning outcome prediction model that explained how to find the suitable prediction model process from various activities on eLearning system. The learning outcome prediction model (LOPM) will help the learner to know the learning result prediction before the end of learning semester. They can increase the learning efficiency and help them to achieve the aims of the study in both knowledge and skills because the systems can record all the information about students' actions and students' interactions, and doing various tasks in log files and database through the learning management.

The learning result prediction vantage not only impacts the learner behavior changing but also it will report a learning situation to the teacher. Then the teacher can change the learning activity to all student or someone specifically. The collection of data from technology-mediated activities to create predictive models of user behavior has been used with increasing consistency in many areas such as marketing, financial markets, sports, health, etc. (Pardo, Han, & Ellis, 2017). There is study that eLearning has brought many advantages to higher education which include:

- eLearning is much less expensive to deliver and conduct than classroom-based education because it does not require any physical plant.
- eLearning is accessible for 24/7 to learners independently from their geographical location.
- eLearning also appeals to the Net Generation's unique needs and expectations in a number of ways.

From the student's perspective, eLearning means increasing opportunities for interaction with other students and instructors and for a wider access to a variety of multimedia resources and experts worldwide (Homiakova et al., 2017).

It can be said that the LOPM presented in this research can be another starting point to increase the potential of the eLearning system that is currently in use. If the eLearning system with the predictive learning tools for learners and instructors is known the learning result before the end of the semester, it will be a great opportunity to make eLearning education effective.

Using LOPM can be developed into eLearning system plug-in that can be added to the eLearning systems such as Moodle. This LOPM also supported the idea of the dynamic eLearning prediction model based on incomplete activities of eLearning system.

This research also presents another important part: a dynamic eLearning model based on the incomplete activities of the eLearning system. This model is an overview model that was created to bring LOPM's performance to the current eLearning system.

It presents the relationship of many elements related to eLearning education, such as learner, instructor, learning prediction, and the element of profit-oriented educational management. By the way, it concerned to the eLearning system value-added and affected to the ROI (Return on Investment) value of eLearning implementation in educational institution as well. These elements have been analyzed to find consistent links, as well as to feedback data between them as shown in Figure 5.1.

The presentation of the model in this overview is to confirm that if the LOPM has been developed in a highly efficient way, the teaching and learning process of higher education institutions will be beneficial in all aspects. On the academic aspect, the process to discovering activities that effect to the learning outcome prediction is the important evidence of data mining technique.

6.4 Limitation and Future work

eLearning has an essential role in the field of modern education. Learning outside the framework of an educational institution and the supervision of a teacher may bring about certain obstacles. Learning in open online platforms requires that students apply self-regulation (Kőrösi & Havasi, 2017). They must manage the time for studying (Wardaya & Pradipto, 2017). With technology in education, people enable them to learn anytime and anywhere. eLearning encourages teachers and students to take personal responsibility for their learning. The data set collecting was focused on an eLearning system that has been used for over three years. There are many courses to choose from the data mining process and many logs gathered by eLearning user behavior. For this study, the data set was collected from one eLearning system due to most of the eLearning system were confidential and restricted. At the same time, some eLearning system never turns on the logs backup function that makes it challenging to find the eLearning logs. So it would be a good idea to get much information from a variety of sources for studying the predictions of learning outcomes based on a large number of activities.

Based on this study, all courses were not chosen by research methodology intention but the suitable courses determined by the machine learning process. Hence, some courses that we wanted to know about and its learning outcome prediction model could not be processed by this study result model. This limitation depended on the course usage behavior logs. Thereby, the eLearning using should encourage eLearning management policy of institution for more complete logs and the better learning outcome prediction model.

This study is a collection of data from all occurrences in the eLearning system. The data mining process was then used to prepare for the screening of sufficient data for use in various procedures. The information used in the processing is delicate and complicated due to the imperfections of the activities that we study. In summary, it is essential to take an in-depth look at every step of data management to obtain the right information to learn about predicting learning outcomes. Most of the models have high predictive accuracy, but at the same time, the value of reliability is low. These causes may be the effect of studying data on variance incompleteness.

The dynamic eLearning prediction model based on incomplete activities of eLearning system is the whole model that proposed the related determinants of eLearning potential improvement. This study focused on the learning outcome prediction model, which is the most crucial component of the whole model. Student performance prediction via online learning behavior analytics can help developers evaluate eLearning system effectively, improve system availability and expand system function. Although we can predict student performance by learning behavior, it is only a part of the learning process. Therefore, other components need to probe deeper into detail for more advantage.

REFERENCES

Al-Alwani, A. (2014). Evaluation Criterion for Quality Assessment of E-Learning Content. *E-Learning and Digital Media*, *11*(6), 532–542. https://doi.org/10.2304/elea.2014.11.6.532

Al Mazidi, A., & Abusham, E. (2018). Study of general education diploma students' performance and prediction in Sultanate of Oman, based on data mining approaches. *International Journal of Engineering Business Management*, *10*, 1847979018807020. https://doi.org/10.1177/1847979018807020

Alharbi, H., & Sandhu, K. (2017). Toward a Model for E-learning Recommender System Adoption: a Pilot Study. In *2017 International Conference on New Trends in Computing Sciences (ICTCS)* (pp. 309–313). https://doi.org/10.1109/ICTCS.2017.57

Alias, U. F., Ahmad, N. B., & Hasan, S. (2017). Mining of E-learning behavior using SOM clustering. In *2017 6th ICT International Student Project Conference (ICT-ISPC)* (pp. 1–4). https://doi.org/10.1109/ICT-ISPC.2017.8075350

Andone, D., Ternauciuc, A., & Vasiu, R. (2017). Using Open Education Tools for a Higher Education Virtual Campus. In *2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT)* (pp. 26–30). https://doi.org/10.1109/ICALT.2017.149

Aydin, C. C., & Tirkes, G. (2010). Open source learning management systems in e-learning and Moodle. In *Education Engineering (EDUCON), 2010 IEEE* (pp. 593–600).

Babic, F., Wagner, J., Jadlovska, S., & Lesko, P. (2010). A logging mechanism for acquisition of real data from different collaborative systems for analytical purposes. In *Applied Machine Intelligence and Informatics (SAMI), 2010 IEEE 8th International Symposium on* (pp. 109–112).

Baesens, B., Mues, C., Martens, D., & Vanthienen, J. (2009). 50 years of datamining and OR: upcoming trends and challenges. *Journal of the Operational Research Society*, *60*.

Bansal, R., Mishra, A., & Singh, S. N. (2017). Mining of educational data for analysing students' overall performance. In *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence* (pp. 495–497). https://doi.org/10.1109/CONFLUENCE.2017.7943202

Baolin, Y., Dujuan, Z., Yi, W., Hai, L., Zhaoli, Z., Jiangbo, S., & Yuegong, L. (2017). Research on personalized learning model under informatization environment. In *Proceedings - 2017 International Symposium on Educational Technology, ISET 2017* (pp. 48–52). https://doi.org/10.1109/ISET.2017.19

Basha, A. D., Umar, I. N., & Abbas, M. (2011). Managing ICT resources in E-learning environment: Challenges, issues and a proposed model. *Proceedings - 3rd International Conference on Computational Intelligence, Communication Systems and Networks, CICSyN 2011*. https://doi.org/10.1109/CICSyN.2011.73

Castillo, G., Gama, J., & Breda, A. M. (2006). An Adaptive Pridictive Model for Student Modeling. *Web-Bassed Education: Personalized Learning Environments*.

Chance, P. (2003). The Study of Learning and Behavior. In *Learning & Behavior* (Fifth, pp. 36–56). Belmont, CA: Phoenix Color Corp.

Chanchary, F. H., Haque, I., & Khalid, S. (2008). Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment. In *2008 Workshop on Knowledge Discovery and Data Mining* (pp. 249–253).

Chang, Y. C., Huang, Y. C., & Chu, C. P. (2009). B(2) model: A browsing behavior model based on High-Level Petri Nets to generate behavioral patterns for e-learning. *Expert Systems with Applications*, *36*(10), 12423–12440. https://doi.org/Doi 10.1016/J.Eswa.2009.04.044

Chatteur, F., Carvalho, L., & Dong, A. (2008). Design for Pedagogy Patterns for E-Learning. In *Advanced Learning Technologies, 2008. ICALT '08. Eighth IEEE International Conference on* (pp. 341–343).

Chen, H., Wu, S., Song, C., & Chen, M. (2009). Research on the Learning Theory of E-learning. In *INC, IMS and IDC, 2009. NCM '09. Fifth International Joint Conference on* (pp. 1185–1187).

Chien Ming, C., Chao Yi, L., Te Yi, C., Bin Shyan, J., & Tsong Wuu, L. (2007). Diagnosis of Students' Online Learning Portfolios. In *Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE '07. 37th Annual* (pp. T3D-17-T3D-22).

Chih Ming, C., Chin Ming, H., Shyuan Yi, C., & Chao Yu, L. (2006). Mining Formative Evaluation Rules Using Learning Portfolios for Web-based Learning Systems. In *Neural Networks, 2006. IJCNN '06. International Joint Conference on* (pp. 2386–2393).

Chih Ping, C., & Yi Chun, C. (2008). A Model for Behavioral Patterns in Web-Based Learning Environments. In *Advanced Information Networking and Applications,* *2008. AINA 2008. 22nd International Conference on* (pp. 181–187).

Chun Xia, Q., Hui Bao, C., Chang Yi, L., & Yue Xing, S. (2010). Design an active e-learning system. In *2nd International Conference on Education Technology and Computer (ICETC), 2010* (Vol. 4, pp. V4-222-V4-225).

Chun Yong, Y., & Qi, L. (2007). Personality Mining System in E-Learning by using Improved Association Rules. In *Machine Learning and Cybernetics, 2007 International Conference on* (Vol. 7, pp. 4130–4134).

Dai, S., & Zhang, P. (2008). A data mining algorithm in distance learning. In *CSCWD 2008. 12th International Conference on Computer Supported Cooperative Work in Design, 2008.* (pp. 1014–1017).

Danubianu, M. (2015). A data preprocessing framework for students' outcome prediction by data mining techniques. In *2015 19th International Conference on System Theory, Control and Computing (ICSTCC)* (pp. 836–841). https://doi.org/10.1109/ICSTCC.2015.7321398

Deena, G., & Raja, K. (2017). A study on knowledge based e-learning in teaching learning process. In *2017 International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET)* (pp. 1–6). https://doi.org/10.1109/ICAMMAET.2017.8186686

Dragoş, S., Săcărea, C., & Şotropa, D. (2017). Investigating educational attractors and life tracks in e-learning environments using formal concept analysis. In *2017 25th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)* (pp. 1–5). https://doi.org/10.23919/SOFTCOM.2017.8115551

Dwivedi, S., & Roshni, V. S. K. (2017). Recommender system for big data in education. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)* (pp. 1–4). https://doi.org/10.1109/ELELTECH.2017.8074993

Elfaki, A. O., Duan, Y., Bachok, R., Du, W., Gapar, M., Johar, M., & Fong, S. (2013). Towards Measuring of E-Learning Usability through User Interface. In *2013 Second IIAI International Conference on Advanced Applied Informatics* (pp. 192–194). https://doi.org/10.1109/IIAI-AAI.2013.17

Eom, S. (2010). Relationships between E-Learning Systems and Learning Outcomes: A Path Analysis Model. In *Advanced Learning Technologies (ICALT), 2010 IEEE 10th International Conference on* (pp. 516–520).

Eryou, W., & Jun, Y. (2008). A New E-learning System: Design and Implementation. In *Computer Science and Software Engineering, 2008 International Conference on* (Vol. 5, pp. 745–748).

Feng-jung, L., & Bai-jiun, S. (2007). Learning Activity-Based E-Learning Material Recommendation System. In *Multimedia Workshops, 2007. ISMW '07. Ninth IEEE International Symposium on* (pp. 343–348).

Feng, T., Qinghua, Z., Zhiyong, G., Jin, D., & Renhou, L. (2007). Personalized Learning Strategies in an intelligent e-Learning Environment. In *Computer Supported Cooperative Work in Design, 2007. CSCWD 2007. 11th International Conference on* (pp. 973–978).

Gang, C. (2010). Design Teacher Educational Technique Training Policies Based on Teacher's Personality Characteristics in E-learning. In *Education Technology* *and Computer Science (ETCS), 2010 Second International Workshop on* (Vol. 2, pp. 307–310).

Gowda, R. S., & Suma, V. (2017). A comparative analysis of traditional education system vs. e-Learning. In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 567–571). https://doi.org/10.1109/ICIMIA.2017.7975524

Graf, S., & List, B. (2005). An evaluation of open source e-learning platforms stressing adaptation issues. In *ICALT 2005. Fifth IEEE International Conference on Advanced Learning Technologies, 2005.* (pp. 163–165).

Gulati, A., Batra, A., Khurana, R., & Tripathi, M. M. (2017). Cognitive learning recommendation system in Indian context. In *2017 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH)* (pp. 1–6). https://doi.org/10.1109/ELELTECH.2017.8074988

Gunathilaka, T. M. A. U., Fernando, M. S. D., & Pasqual, H. (2017). Identification of the learning behavior of the students for education personalization. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)* (pp. 364–370). https://doi.org/10.1109/I-SMAC.2017.8058372

Han, J., & Kamber, K. (2000). *Data mining: Concepts and Techniques*. Simon Fraser University.

Harry N. Boone, J., & Boone, D. A. (2012). Analyzing Likert Data. *Journal of Extension*, *50*(2TOT2).

Haythornthwaite, C., & Andrews, R. (2011). *E-learning Theory and Practice*. Singapore 048763: SAGE Publications Asia-Pacific Pte Ltd.

Homiakova, V., Arras, P., & Kozík, T. (2017). Challenges of using ICT in education. In *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)* (Vol. 2, pp. 1094–1097). https://doi.org/10.1109/IDAACS.2017.8095254

Hsien Tang, L., Chih Hua, W., Chia Feng, L., & Shyan Ming, Y. (2009). Annotating Learning Materials on Moodle LMS. In *ICCTD '09. International Conference on Computer Technology and Development, 2009.* (Vol. 2, pp. 455–459).

Hu, D. (2008). Design and Implementation of E-learning Performance Evaluation System. In *Computer Science and Software Engineering, 2008 International Conference on* (Vol. 5, pp. 376–380).

Huertas, F., & Navarro, A. (2017). Feasibility of a Virtual Campus Decoupled From the Underlying e-Learning Platform: The VCAA Virtual Campus. *IEEE Revista Iberoamericana de Tecnologias Del Aprendizaje*, *12*(4), 173–181. https://doi.org/10.1109/RITA.2017.2776438

Jandrić, P., Knox, J., Macleod, H., & Sinclair, C. (2017). Learning in the age of algorithmic cultures. *E-Learning and Digital Media*, *14*(3), 101–104. https://doi.org/10.1177/2042753017731237

Jili, C., Kebin, H., Feng, W., & Huixia, W. (2009). E-learning Behavior Analysis Based on Fuzzy Clustering. In *Genetic and Evolutionary Computing, 2009. WGEC '09. 3rd International Conference on* (pp. 863–866).

Jing, L., Hailong, M., & Jun, H. (2009). Comparative Study of Open-Source E-Learning Management Platform. In *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on* (pp. 1–4).

Kebin, H., Feimin, L., Ming, Z., Feng, W., & Xiaoshuang, X. (2008). Design and Implement on E-learning Behavior Mine System. In *WGEC '08. Second International Conference on Genetic and Evolutionary Computing, 2008.* (pp. 406–409).

Khalil, M. K., & Elkhider, I. A. (2016). Applying learning theories and instructional design models for effective instruction. *Advances in Physiology Education*, *40*(2), 147–156. https://doi.org/10.1152/advan.00138.2015

Kőrösi, G., & Havasi, F. (2017). Moodle-based data mining potentials of MOOC systems at the University of Szeged. In *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 755–760). https://doi.org/10.23919/MIPRO.2017.7973523

Laksitowening, K. A., Yanuarifiani, A. P., & Wibowo, Y. F. A. (2016). Enhancing e-learning system to support learning style based personalization. In *2016 2nd International Conference on Science in Information Technology (ICSITech)* (pp. 329–333). https://doi.org/10.1109/ICSITech.2016.7852657

Ling, G., Xin, X., & YuanChun, S. (2004). Use Web usage mining to assist online e-learning assessment. In *Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on* (pp. 912–913).

Lingyan, W., Jian, L., Lulu, D., & Pengkun, L. (2010). E-Learning Evaluation System Based on Data Mining. In *2nd International Symposium on Information Engineering and Electronic Commerce (IEEC), 2010* (pp. 1–3).

Lipo, W. (2005). *Support Vector Machines: Theory and Applications*. *Studies in Fuzziness and Soft Computing* (Vol. 177).

Lui, Y. (2005). NGCMS: exploring and supporting effective faculty use. In *Course Management System for Learning: Beyond Accidental Pedagogy* (pp. 131–145). Hershey: Idea Group.

Luo, J., & Chen, Z. (2010). Web Data Mining Based Learners Personalized Information Modeling. In *e-Education, e-Business, e-Management, and e-Learning, 2010. IC4E '10. International Conference on* (pp. 454–459).

Mahboob, T., Irfan, S., & Karamat, A. (2016). A machine learning approach for student assessment in E-learning using Quinlan's C4.5, Naive Bayes and Random Forest algorithms. In *2016 19th International Multi-Topic Conference (INMIC)* (pp. 1–8). https://doi.org/10.1109/INMIC.2016.7840094

Mak, M. K. Y., Ho, G. T. S., & Ting, S. L. (2011). A Financial Data Mining Model for Extracting Customer Behavior. *International Journal of Engineering Business Management*, *3*, 16. https://doi.org/10.5772/50937

Manavoglu, E., Building, T., Pavlov, D., & Giles, C. L. (2003). Probabilistic User Behavior Models. *Work*, 1–18. Retrieved from http://www.computer.org/portal/web/csdl/doi/10.1109/ICDM.2003.1250921

Milani, A., Jasso, J., & Suriani, S. (2008a). Modeling Online User Behavior. In *e-Business Engineering, 2008. ICEBE '08. IEEE International Conference on* (pp. 736–741).

Milani, A., Jasso, J., & Suriani, S. (2008b). Soft User Behavior Modeling in Virtual Environments. In *Convergence and Hybrid Information Technology, 2008. ICCIT '08. Third International Conference on* (Vol. 2, pp. 1182–1187).

Min, J. (2005). Development of an e-learning system for teaching machining technology. In *Proceedings of the 2005 International Conference on Active Media Technology, 2005. (AMT 2005).* (pp. 347–352).

Mingming, Z. (2010). Data Mining and Student e-Learning Profiles. In *E-Business and E-Government (ICEE), 2010 International Conference on* (pp. 5405–5408).

Mon Fong, J., Shian Shyong, T., & Shan Yi, L. (1999). Data types generalization for data mining algorithms. In *Systems, Man, and Cybernetics, 1999. IEEE SMC '99 Conference Proceedings. 1999 IEEE International Conference on* (Vol. 3, pp. 928–933 vol.3).

Moodle.org. (2011). About Moodle. Retrieved from http://docs.moodle.org/20/en/About_Moodle

Morimoto, Y., Ueno, M., Yokoyama, S., & Miyadera, Y. (2007). A SCORM-compliant Learning Management System that Enhances Learning By Managing the Learning Itself. In *Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on* (pp. 122–126).

Mudrák, M. (2017). Analysis and implementation of adaptive course in Moodle. In *2017 15th International Conference on Emerging eLearning Technologies and Applications (ICETA)* (pp. 1–6). https://doi.org/10.1109/ICETA.2017.8102513

Ninomiya, T., Nakayama, K., Shimizu, M., Anma, F., & Okamoto, T. (2007). The Design of e-Learning Environment Oriented for Personalized Adaptability. In *Advanced Learning Technologies, 2007. ICALT 2007. Seventh IEEE International Conference on* (pp. 728–729).

Nkomo, L. M., & Nat, M. (2016). Discovering students use of learning resources with educational data mining. In *2016 HONET-ICT* (pp. 98–102).

https://doi.org/10.1109/HONET.2016.7753428

Nongkhai, L. N., & Kaewkiriya, T. (2015). Framework for e-Leaming recommendation based on index of learning styles model. In *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 587–592). https://doi.org/10.1109/ICITEED.2015.7409015

Nordmann, M., & Neumann, J. (2008). Learning Application Suite Creating and Playing SCORM Compatible Web and Computer Based Training. In *Advanced Learning Technologies, 2008. ICALT '08. Eighth IEEE International Conference on* (pp. 572–573).

Nukoolkit, C., Chansripiboon, P., & Sopitsirikul, S. (2011). Improving university e-Learning with exploratory data analysis and web log mining. In *Computer Science & Education (ICCSE), 2011 6th International Conference on* (pp. 176–179).

Ozkan, S., & Koseler, R. (2009). Multi-dimensional evaluation of E-learning systems in the higher education context: An empirical investigation of a computer literacy course. In *Frontiers in Education Conference, 2009. FIE '09. 39th IEEE* (pp. 1–6).

Paľová, D. (2016). Experience with usage of LMS moodle not only for the educational purposes at the educational institution. In *2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 901–906). https://doi.org/10.1109/MIPRO.2016.7522268

Pan, Z., Xue, J., Gao, Y., Wang, H., & Chen, G. (2018). Revealing the Relations

Between Learning Behaviors and Examination Scores via a Prediction System. In *Proceedings of the 2018 2Nd International Conference on Computer Science and Artificial Intelligence* (pp. 414–419). New York, NY, USA: ACM. https://doi.org/10.1145/3297156.3297195

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining University Student Self-Regulated Learning Indicators and Engagement with Online Learning Events to Predict Academic Performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82–92. https://doi.org/10.1109/TLT.2016.2639508

Peng, M. Y. P., Tuan, S. H., & Liu, F. C. (2017). Establishment of big data application platform for education industry. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)* (pp. 975–978). https://doi.org/10.1109/ICIVC.2017.7984700

Ravichandran, M., & Kulanthaivel, G. (2015). Intelligent prediction model for learners outcome forecasting in e-learning. In *2015 International Conference on Computing and Communications Technologies (ICCCT)* (pp. 7–11). https://doi.org/10.1109/ICCCT2.2015.7292711

Ribeiro, B., & Cardoso, A. (2008). Evaluation system for e-learning with pattern mining tools. In *SMC 2008. IEEE International Conference on Systems, Man and Cybernetics, 2008.* (pp. 3051–3056).

Romero, C., Ventura, S., & Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384. Retrieved from http://www.sciencedirect.com/science/article/B6VCJ-4P7FCT6-1/2/60b47e7bb2ae46aed61cc82a01446798

Ruiz Reyes, N., Vera Candeas, P., Galan, S. G., Viciana, R., Canadas, F., & Reche, P. J. (2009). Comparing open-source e-learning platforms from adaptivity point of view. In *EAEEIE Annual Conference, 2009* (pp. 1–6).

Sanchez-Santillan, M., Paule-Ruiz, Mp., Cerezo, R., & Nuñez, Jc. (2016). Predicting Students' Performance: Incremental Interaction Classifiers. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (pp. 217–220). New York, NY, USA: ACM. https://doi.org/10.1145/2876034.2893418

Sancristobal, E., Castro, M., Harward, J., Baley, P., DeLong, K., & Hardison, J. (2010). Integration view of Web Labs and Learning Management Systems. In *Education Engineering (EDUCON), 2010 IEEE* (pp. 1409–1417).

scorm.com. (2012). Shareable Content Object Reference Model (SCORM). Retrieved from http://scorm.com/scorm-explained/

Sheshasaayee, A., & Bee, M. N. (2017). Evaluating user experience in moodle learning management systems. In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 735–738). https://doi.org/10.1109/ICIMIA.2017.7975562

Sheshasaayee, A., & Malathi, S. (2017). Impact and consequences of BIG DATA in e-learning. In *2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 726–729). https://doi.org/10.1109/ICIMIA.2017.7975560

Solak, E., & Cakir, R. (2015). Language learning strategies of language e-learners in Turkey. *E-Learning and Digital Media*, *12*(1), 107–120. https://doi.org/10.1177/2042753014558384

Susanto, D., Irdoni, S., & Rasyid, M. U. H. A. (2017). Attendance report plugin for E-learning applications in PENS: (Based on moodle). In *2017 International Electronics Symposium on Knowledge Creation and Intelligent Computing (IES-KCIC)* (pp. 153–160). https://doi.org/10.1109/KCIC.2017.8228579

Tan, P. J. B. (2015). English e-learning in the virtual classroom and the factors that influence ESL (English as a Second Language): Taiwanese citizens' acceptance and use of the Modular Object-Oriented Dynamic Learning Environment. *Social Science Information*, *54*(2), 211–228. https://doi.org/10.1177/0539018414566670

Tanasa, D., & Trousse, B. (2004). Advanced data preprocessing for intersites Web usage mining. *Intelligent Systems, IEEE*, *19*(2), 59–65.

Tashtoush, Y. M., Al-Soud, M., Fraihat, M., Al-Sarayrah, W., & Alsmirat, M. A. (2017). Adaptive e-learning web-based English tutor using data mining techniques and Jackson's learning styles. In *2017 8th International Conference on Information and Communication Systems (ICICS)* (pp. 86–91). https://doi.org/10.1109/IACS.2017.7921951

The, M. M., & Usagawa, T. (2017). Evaluation on e-learning readiness of Yangon and Mandalay technological universities, Myanmar. In *TENCON 2017 - 2017 IEEE Region 10 Conference* (pp. 2072–2076). https://doi.org/10.1109/TENCON.2017.8228202

Ting, F. (2008). The Research of E-learning Resources' Construction based on the Discovery-Driven Exploration of Data Cubes. In *Knowledge Acquisition and Modeling Workshop, 2008. KAM Workshop 2008. IEEE International*

Symposium on (pp. 1098–1101).

Tsung Ho, L., Kun Te, W., & Yueh Min, H. (2008). An exploration study on student online learning behavior patterns. In *IT in Medicine and Education, 2008. ITME 2008. IEEE International Symposium on* (pp. 854–859).

Villegas-Ch, W., & Luján-Mora, S. (2017). Analysis of data mining techniques applied to LMS for personalized education. In *2017 IEEE World Engineering Education Conference (EDUNINE)* (pp. 85–89). https://doi.org/10.1109/EDUNINE.2017.7918188

Vivekananthamoorthy, N., Sankar, S., Siva, R., & Sharmila, S. (2009). An effective E-learning framework model - a case study. In *ICT and Knowledge Engineering, 2009 7th International Conference on* (pp. 8–14).

Vyas, M. S., & Gulwani, R. (2017). Predictive analytics for E learning system. In *2017 International Conference on Inventive Systems and Control (ICISC)* (pp. 1–4). https://doi.org/10.1109/ICISC.2017.8068590

Wang, Y., Lu, Y., Zhao, T., & Cha, J. (2010). Research of Data Model of SCORM Run-time Environment. In *Information Management, Innovation Management and Industrial Engineering (ICIII), 2010 International Conference on* (Vol. 2, pp. 240–243).

Wardaya, A., & Pradipto, Y. D. (2017). Factors for improving quality education startup : A review of the literature. In *2017 International Conference on Applied Computer and Communication Technologies (ComCom)* (pp. 1–4). https://doi.org/10.1109/COMCOM.2017.8167088

Wen Hai, G. (2010). Research on client behavior pattern recognition system based on

Web log mining. In *ICMLC 2010 International Conference on Machine Learning and Cybernetics* (Vol. 1, pp. 466–470).

whatislms.com. (2012). What is a Learning Management System. Retrieved from http://www.whatislms.com/what-is-a-learning-management-system/

Witten, I. H., & Frake, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques* (Second). San Francisco, CA: Morgan Kaufmann.

Wu, H. yan, Zhu, J. jun, & Zhang, X. yu. (2009). The Explore of the Web-Based Learning Environment Base on Web Sequential Pattern Mining. In *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on* (pp. 1–6).

Wu, L., Xu, J., & Qu, X. (2009). Exploration of Moodle-Based Collaborative Learning for the Deaf. In *Computer Science-Technology and Applications, 2009. IFCSTA '09. International Forum on* (Vol. 2, pp. 145–147).

Xiaokai, X., Qiuhong, P., Yongpo, L., Ji, W., & Chao, L. (2010). Multi-level logs based web performance evaluation and analysis. In *Computer Application and System Modeling (ICCASM), 2010 International Conference on* (Vol. 4, pp. V4-37-V4-41).

Xu, W., & Jun, Y. (2009). Learner Profile Design for Personalized E-Learning Systems. In *Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on* (pp. 1–4).

Yamada, Y., & Hirakawa, M. (2015). A Case Study of Analyzing Logs of LMS in Flipped Classroom. In *2015 IIAI 4th International Congress on Advanced Applied Informatics* (pp. 374–378). https://doi.org/10.1109/IIAI-AAI.2015.255

Yel, M. B., & Sfenrianto, S. (2017). E-learning model for equivalency education program in Indonesia. In *2017 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)* (pp. 1–5). https://doi.org/10.1109/EECSI.2017.8239198

Yen Hung, K., Juei Nan, C., Yu Lin, J., & Yueh Min, H. (2005). Real-time learning behavior mining for e-learning. In *Web Intelligence, 2005. Proceedings. The 2005 IEEE/WIC/ACM International Conference on* (pp. 653–656). https://doi.org/10.1109/wi.2005.125

Yi, L. y., Zuo, M. z., & Wang, Z. x. (2007). A Model for Analyzing and Evaluating the Return on Investment in e-Learning. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)* (pp. 79–81). https://doi.org/10.1109/ICALT.2007.20

Yongquan, L., Zhongying, Z., & Qingtian, Z. (2007). Mining User's Interest from Reading Behavior in E-learning System. In *Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007. Eighth ACIS International Conference on* (Vol. 2, pp. 417–422).

Zaki, N. M., Deris, S., & Chin, K. K. (2002). Extending the Decomposition Algorithm for Support Vector Machines Training. *Journal of Information and Communication Technology*, *1*, 17–19.

Appendix A

Course Name List

The selected eLearning courses from Moodle log

Course ID	Course Name
101	Electronic Marketing
107	Data Communication and Computer Networks
12	Food Safety and Sanitation
126	Business Computer Seminar
127	Analysis and design of business information systems
128	Computer Operating System
146	Monsters and Characters
149	Human Resource Management
165	Business Communication and Computer Network Management.
178	Graphic Design in Business
190	Processing and Analysis of Business Research Data
191	Introduction to Business Programming
195	Service Psychology
204	Appraisal
206	Introduction to Business
28	Psychology for Development
29	Information System for Marketing Management
33	Data Structures and Algorithms
34	General Physics
36	General Economics

Appendix B

Factor Analysis

Descriptive Statistics			
	Mean	Std. Deviation	Analysis N
assignment_upload	.405307	.3611372	874
assignment_view	.321762	.2845133	874
course_view	.292419	.2161743	874
forum_add_discussion	.021710	.0678019	874
forum_add_post	.305052	.2732581	874
forum_delete_discussion	.004672	.0675478	874
forum_delete_post	.0335	.11765	874
forum_subscribe	.006865	.0473680	874
forum_unsubscribe	.009	.0650	874
forum_update_post	.032545	.0680034	874
forum_view_discussion	.276888	.2672631	874
forum_view_forum	.144027	.1669013	874
quiz_attempt	.055727	.0792751	874
quiz_close_attempt	.054448	.0787379	874
quiz_continue_attempt	.210372	.3165550	874
quiz_review	.016455	.0247408	874
quiz_view	.060553	.0968687	874
quiz_view_summary	.052631	.0822012	874
resource_view	.202204	.1825741	874
url_view	.054511	.1492978	874

Communalities		
	Initial	Extraction
assignment_upload	1.000	.815
assignment_view	1.000	.895
course_view	1.000	.822
forum_add_discussion	1.000	.937
forum_add_post	1.000	.842
forum_delete_discussion	1.000	.937
forum_delete_post	1.000	.607
forum_subscribe	1.000	.890
forum_unsubscribe	1.000	.906
forum_update_post	1.000	.490
forum_view_discussion	1.000	.846
forum_view_forum	1.000	.695
quiz_attempt	1.000	.937
quiz_close_attempt	1.000	.930
quiz_continue_attempt	1.000	.932
quiz_review	1.000	.897
quiz_view	1.000	.938
quiz_view_summary	1.000	.940
resource_view	1.000	.724
url_view	1.000	.518

Extraction Method: Principal Component Analysis.

Total Variance Explained					
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variance	Cumulative %	Total	% of Variance
1	9.982	49.909	49.909	9.982	49.909
2	2.415	12.073	61.983	2.415	12.073
3	2.101	10.506	72.489	2.101	10.506
4	2.001	10.005	82.494	2.001	10.005
5	.789	3.944	86.437		
6	.611	3.054	89.491		
7	.539	2.697	92.188		
8	.393	1.966	94.154		
9	.274	1.370	95.524		
10	.265	1.325	96.849		
11	.160	.800	97.649		
12	.117	.584	98.233		
13	.103	.513	98.746		
14	.085	.427	99.173		
15	.072	.361	99.535		
16	.037	.185	99.719		
17	.027	.133	99.852		
18	.022	.111	99.963		
19	.005	.023	99.986		
20	.003	.014	100.000		

Extraction Method: Principal Component Analysis.

Total Variance Explained				
Component	Extraction Sums of Squared Loadings	Rotation Sums of Squared Loadings		
	Cumulative %	Total	% of Variance	Cumulative %
1	49.909	6.568	32.839	32.839
2	61.983	5.247	26.235	59.074
3	72.489	2.516	12.579	71.653
4	82.494	2.168	10.840	82.494
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				

Extraction Method: Principal Component Analysis.

Component Matrix[a]				
	Component			
	1	2	3	4
assignment_upload	.850	.008	-.245	.182
assignment_view	.843	.024	-.235	.359
course_view	.658	.057	-.121	.609
forum_add_discussion	.290	.833	.364	-.161
forum_add_post	.897	.070	-.127	.130
forum_delete_discussion	.113	.864	.394	-.153
forum_delete_post	.296	.698	.177	-.017
forum_subscribe	.056	-.351	.851	.199
forum_unsubscribe	.060	-.346	.852	.239
forum_update_post	.432	-.222	.346	.366
forum_view_discussion	.880	.088	-.081	.240
forum_view_forum	.752	.069	.130	.329
quiz_attempt	-.931	.109	-.008	.241
quiz_close_attempt	-.923	.112	-.010	.255
quiz_continue_attempt	-.909	.139	-.057	.287
quiz_review	-.899	.131	-.056	.260
quiz_view	-.873	.153	-.051	.388
quiz_view_summary	-.880	.141	-.027	.380
resource_view	.640	.205	-.166	.496
url_view	-.542	.165	-.078	.436

Extraction Method: Principal Component Analysis.[a]

a. 4 components extracted.

Rotated Component Matrix[a]				
	Component			
	1	2	3	4
assignment_upload	-.517	.732	.009	-.112
assignment_view	-.401	.855	.002	-.059
course_view	-.105	.896	.014	.089
forum_add_discussion	-.139	.075	.955	.010
forum_add_post	-.575	.703	.125	-.039
forum_delete_discussion	.005	-.035	.967	.016
forum_delete_post	-.078	.212	.742	-.069
forum_subscribe	-.040	-.059	-.009	.941
forum_unsubscribe	-.017	-.027	-.010	.951
forum_update_post	-.169	.435	-.056	.518
forum_view_discussion	-.493	.763	.140	.025
forum_view_forum	-.356	.695	.174	.236
quiz_attempt	.876	-.404	-.073	-.041
quiz_close_attempt	.879	-.388	-.071	-.039
quiz_continue_attempt	.896	-.342	-.067	-.081
quiz_review	.871	-.356	-.068	-.084
quiz_view	.934	-.245	-.061	-.049
quiz_view_summary	.931	-.262	-.062	-.026
resource_view	-.124	.829	.144	-.036
url_view	.718	.008	-.020	-.041

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.[a]
a. Rotation converged in 5 iterations.

Component Transformation Matrix

Component	1	2	3	4
1	-.755	.637	.143	.068
2	.234	.111	.901	-.347
3	-.044	-.235	.384	.892
4	.612	.726	-.140	.281

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Component Score Coefficient Matrix

	Component			
	1	2	3	4
assignment_upload	-.003	.148	-.043	-.074
assignment_view	.053	.211	-.047	-.047
course_view	.145	.279	-.034	.031
forum_add_discussion	.002	-.042	.393	.014
forum_add_post	-.019	.122	.007	-.039
forum_delete_discussion	.020	-.053	.407	.022
forum_delete_post	.036	.025	.299	-.026
forum_subscribe	.005	-.035	.012	.440
forum_unsubscribe	.017	-.021	.011	.445
forum_update_post	.051	.112	-.039	.233
forum_view_discussion	.017	.156	.014	-.007
forum_view_forum	.048	.156	.038	.097
quiz_attempt	.155	.034	.009	.008
quiz_close_attempt	.159	.040	.009	.009
quiz_continue_attempt	.171	.059	.008	-.010
quiz_review	.162	.049	.008	-.012
quiz_view	.200	.098	.008	.005
quiz_view_summary	.197	.091	.009	.016
resource_view	.126	.249	.021	-.026
url_view	.192	.140	.009	.001

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Component Score Covariance Matrix

Component	1	2	3	4
1	1.000	.000	.000	.000
2	.000	1.000	.000	.000
3	.000	.000	1.000	.000
4	.000	.000	.000	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

Appendix C

Classification Result

Weka Confusion Matrix and Classification Statistics

Activities Group 1: quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review, url_view.

Algorithm: ZeroR

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor1
Instances:    874
Attributes:   8
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor1
Instances:    874
Attributes:   8
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2952
Root mean squared error                  0.3799
Relative absolute error                110.744  %
Root relative squared error            104.1845 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.511     0.845     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.511     0.161     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.511     0.737

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.clas
Relation:     factor1
Instances:    874
Attributes:   8
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor1
Instances:    874
Attributes:   8
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5'
Relation:     factor1
Instances:    874
Attributes:   8
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor1
Instances:    874
Attributes:   8
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         734               83.9817 %
Incorrectly Classified Instances       140               16.0183 %
Kappa statistic                         -0.0045
Mean absolute error                      0.2658
Root mean squared error                  0.3703
Relative absolute error                 99.724  %
Root relative squared error            101.5428 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.997    1.000    0.842      0.997    0.913      -0.021   0.510     0.852     Pass
                 0.000    0.003    0.000      0.000    0.000      -0.021   0.510     0.157     Fail
Weighted Avg.    0.840    0.843    0.709      0.840    0.769      -0.021   0.510     0.742

=== Confusion Matrix ===

   a   b   <-- classified as
 734   2 |   a = Pass
 138   0 |   b = Fail
```

Activities Group 2: course_view, assignment_view, resource_view, forum_view_discussion, assignment_upload, forum_add_post, forum_view_forum.

Algorithm: ZeroR

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor2
Instances:    874
Attributes:   8
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              forum_add_post
              forum_view_discussion
              forum_view_forum
              resource_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor2
Instances:    874
Attributes:   8
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              forum_add_post
              forum_view_discussion
              forum_view_forum
              resource_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         723               82.7231 %
Incorrectly Classified Instances       151               17.2769 %
Kappa statistic                         -0.0185
Mean absolute error                      0.2766
Root mean squared error                  0.3803
Relative absolute error                103.7773 %
Root relative squared error            104.2851 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.981    0.993    0.841      0.981    0.905      -0.033   0.487     0.832     Pass
                 0.007    0.019    0.067      0.007    0.013      -0.033   0.487     0.150     Fail
Weighted Avg.    0.827    0.839    0.718      0.827    0.764      -0.033   0.487     0.725

=== Confusion Matrix ===

   a   b   <-- classified as
 722  14 |   a = Pass
 137   1 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classi
Relation:     factor2
Instances:    874
Attributes:   8
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              forum_add_post
              forum_view_discussion
              forum_view_forum
              resource_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor2
Instances:    874
Attributes:   8
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              forum_add_post
              forum_view_discussion
              forum_view_forum
              resource_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     factor2
Instances:    874
Attributes:   8
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              forum_add_post
              forum_view_discussion
              forum_view_forum
              resource_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor2
Instances:    874
Attributes:   8
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              forum_add_post
              forum_view_discussion
              forum_view_forum
              resource_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         734               83.9817 %
Incorrectly Classified Instances       140               16.0183 %
Kappa statistic                          -0.0045
Mean absolute error                       0.2721
Root mean squared error                   0.3915
Relative absolute error                 102.0737 %
Root relative squared error             107.3548 %
Total Number of Instances               874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.997    1.000    0.842      0.997    0.913      -0.021   0.440     0.820     Pass
                 0.000    0.003    0.000      0.000    0.000      -0.021   0.440     0.135     Fail
Weighted Avg.    0.840    0.843    0.709      0.840    0.769      -0.021   0.440     0.712

=== Confusion Matrix ===

   a   b   <-- classified as
 734   2 |   a = Pass
 138   0 |   b = Fail
```

Activities Group 3: forum_delete_discussion, forum_add_discussion, forum_delete_post.

<u>Algorithm: ZeroR</u>

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor3
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_add_discussion
              forum_delete_discussion
              forum_delete_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor3
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_add_discussion
              forum_delete_discussion
              forum_delete_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2772
Root mean squared error                  0.366
Relative absolute error                104.0101 %
Root relative squared error            100.3819 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.484     0.839     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.484     0.152     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.484     0.731

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.class
Relation:     factor3
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_add_discussion
              forum_delete_discussion
              forum_delete_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor3
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_add_discussion
              forum_delete_discussion
              forum_delete_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     factor3
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_add_discussion
              forum_delete_discussion
              forum_delete_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor3
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_add_discussion
              forum_delete_discussion
              forum_delete_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2663
Root mean squared error                  0.3668
Relative absolute error                 99.9079 %
Root relative squared error            100.5983 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.476     0.837     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.476     0.143     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.476     0.727

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Activities Group 4: forum_unsubscribe, forum_subscribe, forum_update_post**.**

Algorithm: ZeroR

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor4
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_subscribe
              forum_unsubscribe
              forum_update_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor4
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_subscribe
              forum_unsubscribe
              forum_update_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         729               83.4096 %
Incorrectly Classified Instances       145               16.5904 %
Kappa statistic                          0.0036
Mean absolute error                      0.2664
Root mean squared error                  0.3723
Relative absolute error                 99.9582 %
Root relative squared error            102.0974 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.988    0.986    0.842      0.988    0.909      0.007    0.522     0.844     Pass
                 0.014    0.012    0.182      0.014    0.027      0.007    0.522     0.173     Fail
Weighted Avg.    0.834    0.832    0.738      0.834    0.770      0.007    0.522     0.738

=== Confusion Matrix ===

   a   b   <-- classified as
 727   9 |   a = Pass
 136   2 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.class
Relation:     factor4
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_subscribe
              forum_unsubscribe
              forum_update_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor4
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_subscribe
              forum_unsubscribe
              forum_update_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     factor4
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_subscribe
              forum_unsubscribe
              forum_update_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor4
Instances:    874
Attributes:   4
              learningResult_2_class
              forum_subscribe
              forum_unsubscribe
              forum_update_post
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2667
Root mean squared error                  0.3671
Relative absolute error                100.066  %
Root relative squared error            100.666  %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.836     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.728

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Activities Group 5: forum_add_post, forum_view_forum, forum_view_discussion, forum_delete_discussion, forum_add_discussion, forum_delete_post, forum_unsubscribe, forum_subscribe, forum_update_post.

Algorithm: ZeroR

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor-name1
Instances:    874
Attributes:   10
              learningResult_2_class
              forum_add_discussion
              forum_add_post
              forum_delete_discussion
              forum_delete_post
              forum_subscribe
              forum_unsubscribe
              forum_update_post
              forum_view_discussion
              forum_view_forum
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor-name1
Instances:    874
Attributes:   10
              learningResult_2_class
              forum_add_discussion
              forum_add_post
              forum_delete_discussion
              forum_delete_post
              forum_subscribe
              forum_unsubscribe
              forum_update_post
              forum_view_discussion
              forum_view_forum
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         722               82.6087 %
Incorrectly Classified Instances       152               17.3913 %
Kappa statistic                         -0.0021
Mean absolute error                      0.2793
Root mean squared error                  0.3811
Relative absolute error                104.806  %
Root relative squared error            104.5149 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.977    0.978    0.842      0.977    0.904      -0.003   0.513     0.847     Pass
                 0.022    0.023    0.150      0.022    0.038      -0.003   0.513     0.164     Fail
Weighted Avg.    0.826    0.827    0.733      0.826    0.768      -0.003   0.513     0.739

=== Confusion Matrix ===

   a   b   <-- classified as
 719  17 |   a = Pass
 135   3 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classi:
Relation:     factor-name1
Instances:    874
Attributes:   10
              learningResult_2_class
              forum_add_discussion
              forum_add_post
              forum_delete_discussion
              forum_delete_post
              forum_subscribe
              forum_unsubscribe
              forum_update_post
              forum_view_discussion
              forum_view_forum
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor-name1
Instances:    874
Attributes:   10
              learningResult_2_class
              forum_add_discussion
              forum_add_post
              forum_delete_discussion
              forum_delete_post
              forum_subscribe
              forum_unsubscribe
              forum_update_post
              forum_view_discussion
              forum_view_forum
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     factor-name1
Instances:    874
Attributes:   10
              learningResult_2_class
              forum_add_discussion
              forum_add_post
              forum_delete_discussion
              forum_delete_post
              forum_subscribe
              forum_unsubscribe
              forum_update_post
              forum_view_discussion
              forum_view_forum
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor-name1
Instances:    874
Attributes:   10
              learningResult_2_class
              forum_add_discussion
              forum_add_post
              forum_delete_discussion
              forum_delete_post
              forum_subscribe
              forum_unsubscribe
              forum_update_post
              forum_view_discussion
              forum_view_forum
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         734               83.9817 %
Incorrectly Classified Instances       140               16.0183 %
Kappa statistic                         -0.0045
Mean absolute error                      0.2722
Root mean squared error                  0.388
Relative absolute error                102.1423 %
Root relative squared error            106.3935 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.997    1.000    0.842      0.997    0.913      -0.021   0.427     0.813     Pass
                 0.000    0.003    0.000      0.000    0.000      -0.021   0.427     0.133     Fail
Weighted Avg.    0.840    0.843    0.709      0.840    0.769      -0.021   0.427     0.705

=== Confusion Matrix ===

   a   b   <-- classified as
 734   2 |   a = Pass
 138   0 |   b = Fail
```

Activities Group 6: quiz_view, quiz_view_summary, quiz_continue_attempt, quiz_close_attempt, quiz_attempt, quiz_review.

Algorithm: ZeroR

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor-name2
Instances:    874
Attributes:   7
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor-name2
Instances:    874
Attributes:   7
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2949
Root mean squared error                  0.3791
Relative absolute error                110.6332 %
Root relative squared error            103.9508 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.517     0.851     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.517     0.163     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.517     0.743

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classi:
Relation:     factor-name2
Instances:    874
Attributes:   7
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor-name2
Instances:    874
Attributes:   7
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     factor-name2
Instances:    874
Attributes:   7
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor-name2
Instances:    874
Attributes:   7
              learningResult_2_class
              quiz_attempt
              quiz_close_attempt
              quiz_continue_attempt
              quiz_review
              quiz_view
              quiz_view_summary
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         735               84.0961 %
Incorrectly Classified Instances       139               15.9039 %
Kappa statistic                         -0.0023
Mean absolute error                      0.2649
Root mean squared error                  0.3688
Relative absolute error                 99.3794 %
Root relative squared error            101.1347 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.999    1.000    0.842      0.999    0.914      -0.015   0.504     0.851     Pass
                 0.000    0.001    0.000      0.000    0.000      -0.015   0.504     0.156     Fail
Weighted Avg.    0.841    0.842    0.709      0.841    0.769      -0.015   0.504     0.741

=== Confusion Matrix ===

   a   b   <-- classified as
 735   1 |   a = Pass
 138   0 |   b = Fail
```

Activities Group 7: course_view, assignment_view, assignment_upload, resource_view, url_view.

<u>Algorithm: ZeroR</u>

```
=== Run information ===

Scheme:       weka.classifiers.rules.ZeroR
Relation:     factor-name3
Instances:    874
Attributes:   6
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              resource_view
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2665
Root mean squared error                  0.3646
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: Naive Bayes

```
=== Run information ===

Scheme:       weka.classifiers.bayes.NaiveBayes
Relation:     factor-name3
Instances:    874
Attributes:   6
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              resource_view
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         721               82.4943 %
Incorrectly Classified Instances       153               17.5057 %
Kappa statistic                         -0.0319
Mean absolute error                      0.2755
Root mean squared error                  0.3799
Relative absolute error                103.3644 %
Root relative squared error            104.173  %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC     ROC Area  PRC Area  Class
                 0.980    1.000    0.839      0.980    0.904      -0.057  0.462     0.826     Pass
                 0.000    0.020    0.000      0.000    0.000      -0.057  0.462     0.143     Fail
Weighted Avg.    0.825    0.845    0.707      0.825    0.761      -0.057  0.462     0.718

=== Confusion Matrix ===

   a   b   <-- classified as
 721  15 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: SVM

```
=== Run information ===

Scheme:       weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.clas:
Relation:     factor-name3
Instances:    874
Attributes:   6
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              resource_view
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.1579
Root mean squared error                  0.3974
Relative absolute error                 59.2419 %
Root relative squared error            108.9704 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.500     0.842     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.500     0.158     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.500     0.734

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: J48

```
=== Run information ===

Scheme:       weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:     factor-name3
Instances:    874
Attributes:   6
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              resource_view
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2659
Root mean squared error                  0.3646
Relative absolute error                 99.7773 %
Root relative squared error             99.9997 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: DecisionTable

```
=== Run information ===

Scheme:       weka.classifiers.rules.DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
Relation:     factor-name3
Instances:    874
Attributes:   6
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              resource_view
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         736               84.2105 %
Incorrectly Classified Instances       138               15.7895 %
Kappa statistic                          0
Mean absolute error                      0.2666
Root mean squared error                  0.3647
Relative absolute error                100.019  %
Root relative squared error            100.0002 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    1.000    0.842      1.000    0.914      ?        0.493     0.840     Pass
                 0.000    0.000    ?          0.000    ?          ?        0.493     0.155     Fail
Weighted Avg.    0.842    0.842    ?          0.842    ?          ?        0.493     0.732

=== Confusion Matrix ===

   a   b   <-- classified as
 736   0 |   a = Pass
 138   0 |   b = Fail
```

Algorithm: RandomTree

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1
Relation:     factor-name3
Instances:    874
Attributes:   6
              learningResult_2_class
              assignment_upload
              assignment_view
              course_view
              resource_view
              url_view
Test mode:    10-fold cross-validation

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         734               83.9817 %
Incorrectly Classified Instances       140               16.0183 %
Kappa statistic                         -0.0045
Mean absolute error                      0.2721
Root mean squared error                  0.3915
Relative absolute error                102.0737 %
Root relative squared error            107.3548 %
Total Number of Instances              874

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.997    1.000    0.842      0.997    0.913      -0.021   0.440     0.820     Pass
                 0.000    0.003    0.000      0.000    0.000      -0.021   0.440     0.135     Fail
Weighted Avg.    0.840    0.843    0.709      0.840    0.769      -0.021   0.440     0.712

=== Confusion Matrix ===

   a   b   <-- classified as
 734   2 |   a = Pass
 138   0 |   b = Fail
```

Appendix D

Eight Reviewing Expert

Chosen Eight Reviewing Expert

No.	Name of Expert
1	Asst. Prof. Dr. Chairat Jussapalo Rajamangala University of Technology Srivijaya Songkhla Province, Thailand
2	Asst. Prof. Dr. Phatchakorn Areekul Rajamangala University of Technology Srivijaya Trang Province, Thailand
3	Dr. Kanokwan Watkins Didyasarin International College, Hatyai University Songkhla Province, Thailand
4	Dr. Uearee Juntorn (Lecturer) Suan Dusit University Bangkok, Thailand
5	Dr. Ubonrat Harinwan (Lecturer) Suan Dusit University Lampang Province, Thailand
6	Saowatarn Samanit (Lecturer) Head of Airline Business Program Suan Dusit University Lampang Province, Thailand
7	Asst. Prof. Phuripoj Kaewyong (Lecturer) Suan Dusit University Bangkok, Thailand
8	Asst. Prof. Dr.Surasit Songma (Lecturer) Suan Dusit University Bangkok, Thailand

Appendix E

Expert Review

Expert Review Document

Research Information

This expert review document for the eLearning expert's point of view for the result of the research title "A Dynamic eLearning Prediction Model Based on Incomplete Activities of eLearning System". This expert opinion is part of the PhD research (Ph.D. Information Technology) of Awang Had Salleh Graduate School of Arts and Science (AHSGS), Universiti Utara Malaysia.

Research Title:

A Dynamic eLearning Prediction Model Based on Incomplete Activities of eLearning System.

Researcher Name:

Mr.Songsakda Chayanukro

Research Objective:

1. To analyze the eLearning activities that affect learning outcome.
2. To construct a learning outcome prediction model for eLearning usage.
3. To synthesize a dynamic eLearning prediction model based on incomplete activities of eLearning systems.
4. To evaluate the dynamic eLearning prediction model based on incomplete activities of an eLearning system on advantage, accuracy, and effectiveness.

Research Ideas:

This research aimed to find out how to improve the efficiency of teaching and learning by eLearning in higher educational based on eLearning using truly condition. The type of eLearning using levels such as eLearning replacing all classroom instruction or use eLearning together with traditional classroom instruction or use it in some courses. The not complete eLearning system using is used in most institutions. The eLearning using condition depended on the policy or the concept of teaching and learning management. This makes the use of eLearning system at the institutes are not consistent. Therefore, finding a machine is very effective in helping learners learn eLearning habits of learners that can affect the learning outcomes. It requires a highly effective prediction tool in order to obtain a model for predicting that activities behavior. Apart from the point of learning to find effective models for predicting learning behavior. This work also focuses on the issue of the achievement of the eLearning system as a tool for teaching and learning as an investment in education.

This will be an opportunity for the university to reduce the cost of teaching and learning, while enhancing its learning ability.

In addition to the issues of learning to find a model that is effective in predicting learning behavior. This research also focuses on the achievement of the implementation of eLearning system to help in teaching and learning. In view of the investment in education. It is also an opportunity to reduce the cost of teaching and learning. At the same time, it enhances the learner's ability to learn. The conceptual framework of research is as follows.

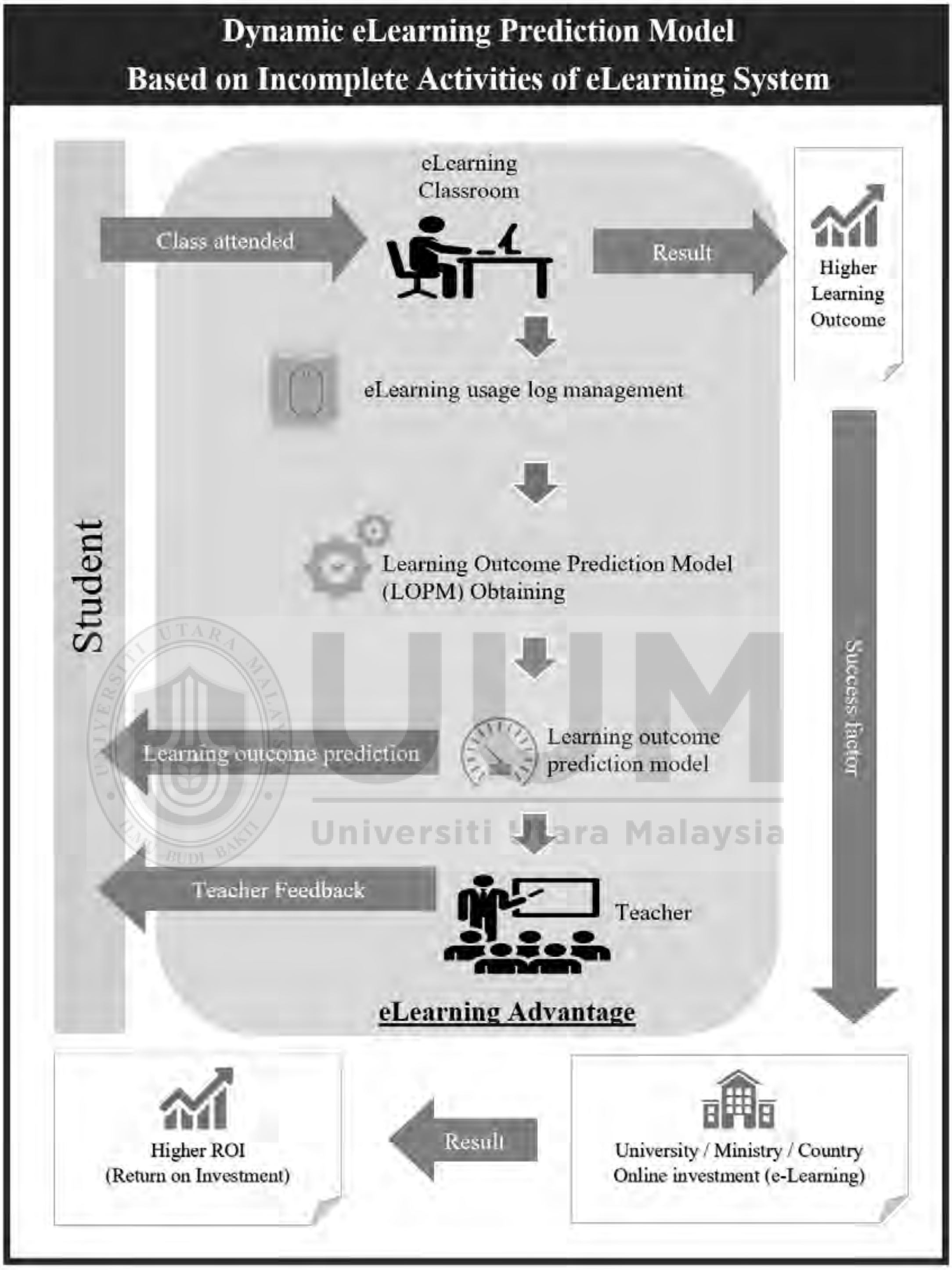

Figure 1. A Dynamic eLearning prediction model based on incomplete activities of eLearning system

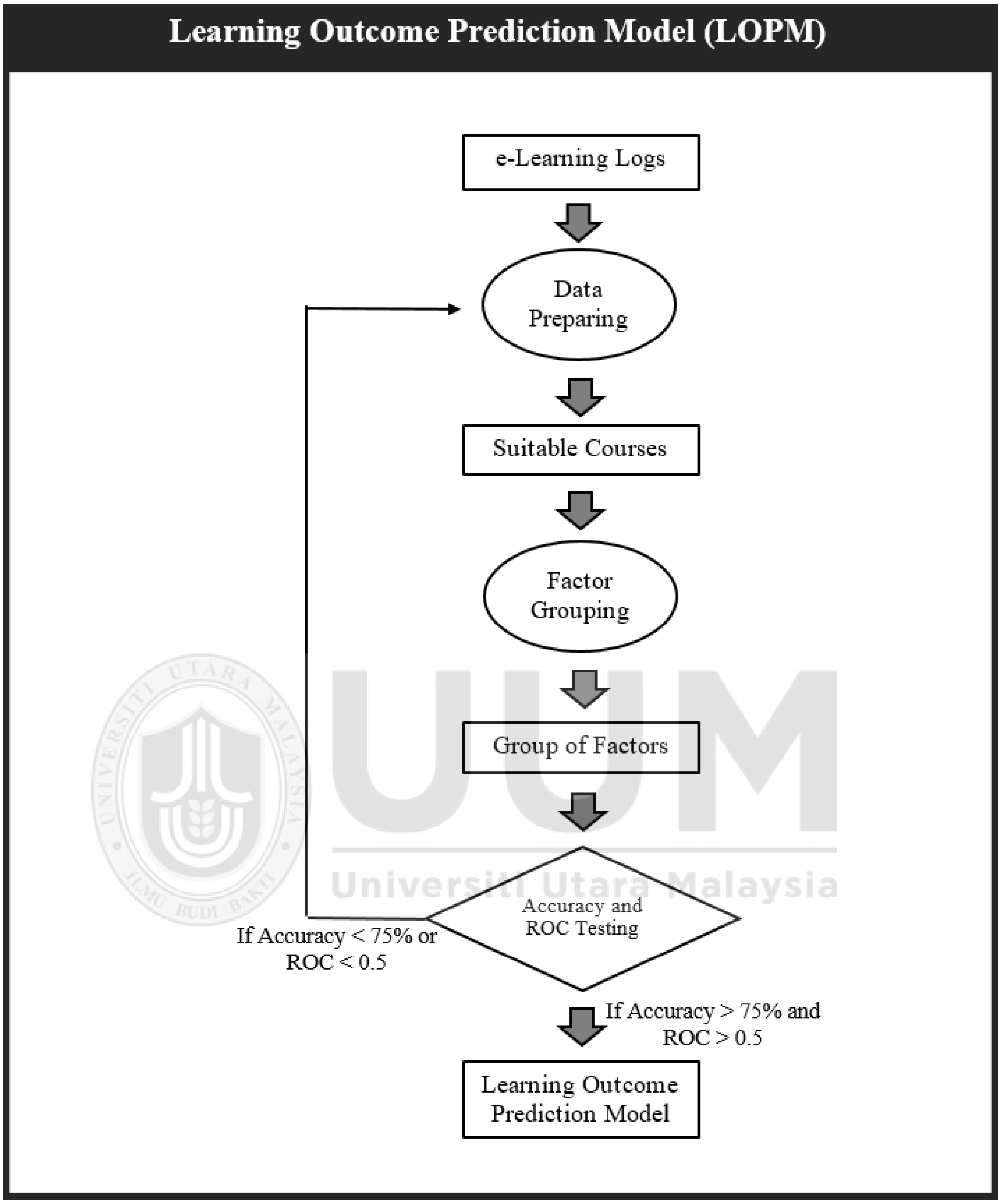

Figure 2. Learning Outcome Prediction Model (LOPM)

From the conceptual framework of the above model, it is explained that once the eLearning system has been integrated into the teaching process, History of the system (log) can be used to study the learning behavior of students that affect the learning. The process of preparing the log and the process of identifying the activities that affect learning will screen the activities sufficiently and import them into the data mining process. Then, the activities were entered into relationship analysis by machine learning software. In this step, the result is a learning outcome prediction model. Based on this historical data, the model can be modeled as seven predictive models from seven activities groups as shown in the following table.

Table: Learning outcome prediction models performance ranking

Output model	ROC	Accuracy Ratio	Ranking
Model 1	0.510	0.839	1
Model 2	0.440	0.839	5
Model 3	0.476	0.842	4
Model 4	0.493	0.842	3
Model 5	0.427	0.839	6
Model 6	0.504	0.841	2
Model 7	0.440	0.839	5

Based on the results of the seven models, the efficiency of the model can be described using the Accuracy Ratio and ROC criteria.

- Accuracy Ratio will indicate the correctness to predict the results of the model. If the predictive value of the sample set is greater than 75%, then the model is acceptable. From the table, it is known that all seven models have a value of 0.75. It is concluded that all models have accuracy in predicting higher than the standard.
- The ROC will indicate the predictive efficiency of the model. If the ROC is higher than 0.5 , then the model is at the low discrimination level. From the table, we know that there is a model number 1)model 1) and model number 6)model 6) get the value of 0.5. It can be concluded that these two models are highly effective in standardization.

If the criteria of the accuracy ratio and ROC are used, the first model and the sixth model are models with high predictive validity and model efficiency at low discrimination level. It can be seen that the predictive modeling process may receive more than one reliable model. It is good to be able to choose the predictive model that best suits the activities that appear in each of the individual subjects.

From the model above, it can be explained that if the modeling process above can be modeled, it will be used to predict the learner's performance accurately and effectively. We can predict students' grades before the end of the semester. This is a guideline for modifying the learner's behavior. Also, the teacher's teaching behavior can be changed through the predictive effect. Predictable learning outcomes and behavioral modifications can affect the educational outcomes of higher education institutions. It is a success activities of eLearning system to help in the process of teaching and learning units of education. In terms of investment in education, it may be a higher return on investment (ROI).

Expert Personal Information

Name:

Education:

Academic Position:

Work Position:

How long have you been involved with eLearning (year)?

What the activities do you work with eLearning?

Expert Review Question

Question 1:

<table>
<tr><th>Feedback / Question</th><th colspan="5">The components of the model are complete, which can be explained the modeling, predicting learning outcomes and the advantage.</th></tr>
<tr><td>Agreement Level</td><td>☐
Strongly Disagree</td><td>☐
Disagree</td><td>☐
Neutral</td><td>☐
Agree</td><td>☐
Strongly Agree</td></tr>
<tr><td>Expert Reviewing</td><td colspan="5"></td></tr>
</table>

Question 2:

Feedback / Question	**"A dynamic eLearning prediction model based on incomplete activities of eLearning system" is accurate in the process of developing the learning outcome prediction model.**				
Agreement Level	☐ Strongly Disagree	☐ Disagree	☐ Neutral	☐ Agree	☐ Strongly Agree
Expert Reviewing					

Question 3:

Feedback / Question	**This "Learning Outcome Prediction Model" effective enough to predict the learning outcomes.**				
Agreement Level	☐ Strongly Disagree	☐ Disagree	☐ Neutral	☐ Agree	☐ Strongly Agree
Expert Reviewing					

Question 4:

Feedback / Question	What should be added to the "A Dynamic eLearning prediction model based on incomplete activities of eLearning system" (eg, model development, performance measurement, model impact study)?
Expert Reviewing	