MODELLING FACTORS AFFECTING STUDENTS' AND INSTRUCTORS' BEHAVIOR AND ENGAGEMENT IN THE ONLINE ENVIRONMENT USING EDUCATIONAL DATA MINING AND DEMATEL TECHNIQUE

HANAN AHMAD YAHYA ALDOWAH

UNIVERSITI SAINS MALAYSIA

2023

MODELLING FACTORS AFFECTING STUDENTS’ AND INSTRUCTORS’ BEHAVIOR AND ENGAGEMENT IN THE ONLINE ENVIRONMENT USING EDUCATIONAL DATA MINING AND DEMATEL TECHNIQUE

by

HANAN AHMAD YAHYA ALDOWAH

Thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

June 2023

ACKNOWLEDGEMENT

First and foremost, thanks, praise, and gratitude to Allah, who bestowed me with good health, sound mind and conducive facilities to conduct and document this research. Secondly, the completion of this research and my PhD programme as a whole would not have been possible without the support and guidance of so many great people. Without their help, I would not have been able to finish this long and very interesting journey. Firstly, I would like to express my deep gratitude to my main supervisor Prof Dr. Irfan Naufal Umar for his efforts, support, guidance, patience and the many words of encouragement through this journey. I would like to extend my appreciation and thanks to my co-supervisors for their valuable guidance. My special gratitude goes to my family, especially my mother, brothers, and sisters for their kind cooperation, support, encouragements and love which helped me in complementing this journey. I am especially indebted to my mother and my brother Mohammed Aldowah for the generous support they have given me. There are not enough words to express all of the love and appreciation that I feel for all of them for the encouragement they have given me throughout this journey. This thesis is also dedicated to the soul of my father; may Allah grant him the highest paradise. I am especially grateful to all my friends who were with me throughout this journey. I would like to extend my appreciation and thanks to the Director of the Center for Instructional Technology and Multimedia and to all the academics, and the administrative staffs at the center for the great support and facilities that have provided to make this journey easy and enjoyable. Last but not least, I express my special gratitude to the Republic of Yemen and USM for facilitating the process of conducting this study and the financial support that enabled me to complete my postgraduates' studies.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

LIST OF SYMBOLS

A	Direct relation matrix
C	Sum of column (i)
c	Cluster centroid
D	Normalized matrix
F	Factor (F=F1, F2, F3,…..,F_n)
I	Identity matrix
i	Number of rows (i=1,2,3.....,n)
j	Number of columns (j=1,2,3.....,n)
K	Number of clusters
N	Total number of participants
n	Maximum number of participants
k	Cohen's kappa statistical coefficient
R	Sum of row (i)
S	Maximum scalar between rows and columns
T	Total relation matrix
X	Data point in a cluster

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
CoI	Community of Inquiry
CRISP-DM	CRoss Industry Standard Process for Data Mining
CSCL	Computer Supported Collaborative Learning
CSEQ	College Student Experiences Questionnaire
DBI	Davies–Bouldin Index
DEMATEL	Decision Making Trial and Evaluation Laboratory
DM	Data Mining
EDM	Educational Data Mining
GUI	Graphical User Interface
HEI	Higher Education Institutions
KDD	knowledge Discovery in Databases
LA	Learning Analytics
LMS	Learning Management System
ML	Machine Learning
Moodle	Modular Object-Oriented Dynamic Learning Environment
NSSE	National Survey of Student Engagement
SAS	Statistical Analysis Systems
SEMMA	Sample, Explore, Modify, Model, Assess
SEQ	Student Engagement Questionnaire
SLR	Systematic Literature Review
USM	Universiti Sains Malaysia

LIST OF APPENDICES

PEMODELAN FAKTOR-FAKTOR PERLAKUAN DAN PENGLIBATAN PELAJAR DAN INSTRUKTOR DI DALAM PERSEKITARAN DALAM TALIAN MENGGUNAKAN PERLOMBONGAN DATA PENDIDIKAN DAN TEKNIK DEMATEL

ABSTRAK

Sistem pengurusan pembelajaran (LMS) kini digunakan secara meluas bagi menyokong pengajaran dan pembelajaran di peringkat pendidikan tinggi. Platform ini menawarkan maklumat penting berkenaan penggunaan dan perlakuan pengguna dalam persekitaran pembelajaran dalam talian. Oleh yang demikian, penilaian persekitaran LMS adalah penting bagi memaksimumkan keberkesanannya. Tujuan utama kajian ini adalah bagi menyiasat penggunaan LMS dalam kalangan pelajar dan instruktor bagi menilai aktiviti dalam talian, menyiasat perlakuan dan pola penglibatan, serta mendapatkan indikator bagi memahami tahap penglibatan mereka. Bagi tujuan ini, kaedah statistik, visualisasi dan kelompokan dari teknik *Educational Data Mining* (EDM) telah digunakan bagi menganalisis kursus-kursus yang ditawarkan secara pembelajaran hibrid dan secara dalam talian sepenuhnya selama empat semester di Universiti Sains Malaysia. Memandangkan penglibatan pelajar adalah penting bagi mempastikan keberkesanan pembelajaran dalam talian, kajian ini juga berhasrat menyiasat hubungan sebab-musabab antara beberapa faktor yang telah dikenalpasti melalui sorotan kajian lampau serta mengkaji punca tahap penglibatan pelajar yang rendah. Teknik DEMATEL telah digunakan sebagai kaedah yang berkesan bagi mengkaji hubungan sebab-musabab antara faktor-faktor penglibatan dalam talian dan memvisualisasikan struktur ini melalui peta perhubungan sebab-akibat. Pengutipan data dalam kajian ini membabitkan dua instrumen: (i) data log LMS

bagi menilai pola penggunaan dan tahap penglibatan pelajar dan instruktor dalam kursus-kursus dalam talian yang berdasarkan kepada 13 pemboleh ubah utama yang mewakili aktiviti LMS dan alatan umum LMS, dan (ii) temu bual bagi menganggar kekuatan perhubungan sebab-musabab antara faktor-faktor penglibatan dari persepsi pelajar dan instruktor. Sebelum temu bual dijalankan, satu semakan sistematik terhadap kajian-kajian lampau telah dijalankan bagi mengenalpasti faktor-faktor penting yang mempengaruhi penglibatan pelajar. Berdasarkan semakan tersebut, sebanyak 15 faktor telah dikenalpasti dan digunakan bagi menghasilkan matriks DEMATEL untuk digunakan dalam temu bual berkenaan. Berdasarkan analisis kaedah perlombongan data, kajian ini mendapati bahawa penggunaan LMS adalah rendah semasa pelaksanaan semester teradun, dan relatif tinggi semasa semester dalam talian akibat pandemik Covid-19. Juga, tiada perbezaan signifikan antara dua tahun penggunaan LMS tersebut, iaitu dari aspek jenis peralatan LMS dan aktiviti yang dijalankan. Kajian ini juga mendapati bahawa alatan penilaian LMS dan sumber adalah dua komponen yang paling banyak digunakan instruktor dan pelajar, manakala alatan komunikasi dan kolaborasi adalah yang paling kurang digunakan. Seterusnya, berdasarkan analisis kluster terhadap data penglibatan pelajar dalam talian, sebanyak 120 kursus telah dikelompokkan mengikut tahap penglibatan mereka. Tambahan pula, berdasarkan dapatan analisis teknik DEMATEL, lima faktor iaitu pengalaman lampau, reka bentuk kursus, isi kandungan kursus, sokongan pihak universiti, dan struktur dan antara muka LMS telah dikenalpasti oleh pelajar dan instruktor sebagai faktor-faktor utama yang mempengaruhi penglibatan pelajar. Juga, kedua-dua pihak bersetuju bahawa (i) pembelajaran aktif dan kolaboratif, (ii) interaksi pelajar – isi kandungan, (iii) masa untuk tugasan, dan (iv) motivasi adalah faktor-faktor yang kurang penting. Sebaliknya, mereka memberikan pandangan berbeza dari aspek tahap kepentingan faktor-faktor berkaitan aktiviti kursus, kemahiran dan kebolehan akademik, interaksi pelajar-pelajar, interaksi pelajar-instruktor, interaksi sistem, dan maklumbalas. Sebagai rumusan, dapatan-dapatan kajian ini diharapkan dapat membantu pentadbir universiti untuk memahami status terkini penggunaan LMS dan tahap penglibatan pelajar dan instruktor, dan seterusnya mencadangkan strategi tertentu untuk tujuan peningkatan kualiti proses pengajaran dan pembelajaran dalam talian.

MODELLING FACTORS AFFECTING STUDENTS' AND INSTRUCTORS' BEHAVIOR AND ENGAGEMENT IN THE ONLINE ENVIRONMENT USING EDUCATIONAL DATA MINING AND DEMATEL TECHNIQUE

ABSTRACT

Learning Management System or LMS is widely used to support teaching and learning in higher learning institutions. This platform offers valuable information about the users' usage data and behavior in the online environment. Thus, the evaluation of such platform is necessary to maximize its effectiveness. The first aim of this study is to investigate the use of LMS by students and instructors to evaluate their online activities, discover user behavior and engagement patterns, and to obtain some indicators to better understand their level of engagement. For this purpose, statistical, visualization, and clustering Educational Data Mining techniques are used to analyze the courses conducted through hybrid and fully online learning modes offered in four semesters at the Universiti Sains Malaysia. Since students' online engagement is critical for effective online learning, this study also aims to study the causal relationships between a variety of factors reported in the literature that have influenced students' engagement in online environment as well as to investigate the causes of low engagement. The DEMATEL technique was used as an effective method to study the causal relationships between those factors and visualizes this structure by cause–effect relationship map. Two instruments were employed in this study: (1) the LMS logs data to evaluate students' and instructors' usage pattern and engagement level in the online courses based on 13 main variables representing the common LMS tools and related activities, and (2) the interview to estimate the strength of the causal relationships between online engagement factors form as perceived by them. Prior to the interview, a systematic literature review was first conducted to collect the important factors influencing students' online engagement, in which a list of 15 factors were identified and used to build the DEMATEL matrix for the interview structure. Based on the data mining analysis, the study found that the usage level of LMS was low during the blended semesters and relatively high during the online semesters because of the Covid-19 pandemic. Also, no significant difference was found during the two years in terms of the type of LMS tools used and activities. The study also found that resources and LMS assessment tools were the most used by instructors and students whereas communication and collaboration tools were the least used. Additionally, based on the cluster analysis on the students' engagement, 120 LMS courses were grouped according to their engagement level. Furthermore, based on the findings of the DEMATEL technique, five factors, namely, prior experience, course design, course content, university support, and LMS structure and interface were identified by the instructors and students as the most important factors affecting students' online engagement. Also, both parties agreed that (1) active and collaborative learning, (2) student-content interaction, (3) time on task, and motivation are all less important factors. On the contrary, instructors and students reported different views in terms of the degree of importance of factors such course activities, academic skills and abilities, instructor-student interaction, student-student interaction, system interaction, and feedback. To conclude, the findings of this study are expected to help the university administrators understand the current status of the LMS usage and the level of engagement by both parties, and to identify the necessary strategies to improve the online teaching and learning activities.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Nowadays, higher education institutions (HEI) around the world are undergoing rapid changes as they adapt to the new realities of the knowledge society (Seraji et al., 2022). These HEI, in particular universities, encounter several problems that keep them away from achieving their quality targets. The majority of these problems started from a knowledge gap. The knowledge gap is the lack of important knowledge at the educational main processes such as planning, evaluation, and marketing because most of these institutions are not able to access to the necessary information to provide appropriate recommendation for their students (Kordrostami & Seitz, 2022). Therefore, universities and other HEI must be prepared through the continuous incorporation of new technologies and the widespread adoption of online learning which has become a global phenomenon to provide professional learning in accordance with the required qualifications (Al-Fraihat, Joy, & Sinclair, 2020).

Research suggested that data-driven decision-making improves the productivity of the educational institutions (Varela et al., 2019). Hence, the most dramatic aspect for shaping the future of higher education is big data and data analytics such as educational data mining, learning analytics, academic analytics, visual analytics, etc. (Lesjak et al., 2021; Raju et al., 2020). The ability of these advanced analytics and computational methods to analyze data of human behaviors and cognitive abilities is valuable for managing educational data, as these techniques are able to discover patterns, gaps, clusters, or trends, and derive useful knowledge (Mense et al., 2020). Thus, making decisions based on such knowledge and evidence seems clear (Viloria et al., 2019).

Over the last decade, colleges and universities have witnessed a rapid deployment of LMSs, and the number of courses is being offered online in a growing particularly after Covid-19 pandemic (Bradley, 2021). The implementation of LMS systems such as Moodle and the interactions within these systems generate a wide range of data that many researchers have positively linked to students' performance and learning (Rajabalee, Santally, & Rennie, 2020). Prior studies also emphasized that HEI can extract meaningful information from LMS-generated student tracking data to inform instructors and other decision-makers about student usage behaviors of LMS and student engagement such as communicating with peers, participating in discussion forums, and performing online tasks and tests (Al-Sharhan et al., 2020; Avcı & Ergün, 2022; Ismail et al., 2021; Martinez-Abad & Chaparro-Caso-López, 2017). Therefore, many universities use LMS as an information resource to support on- and off-campus online learning, including support for blended learning, e-learning, as well as face-to-face learning (Ghilay, 2019). Actually, evaluating instructors' and students' usage behaviors and engagement patterns in such case is critical, considering the fact that the current LMS system does not provide much information about the level of student engagement (Henrie et al., 2018; Lee et al., 2019). Educational data mining (EDM) is an emerging discipline that develops methods and techniques to explore unique and extensive educational data and extract useful information as well as create predictive models that help to improve learning and teaching processes (Martínez-Abad et al., 2020). According to Hernández-Blanco et al., (2019), EDM can be very useful in discovering hidden patterns and valuable information that can be used to characterize students based on their academic records to understand how students engage and interact in the settings in which they learn. Meanwhile, Chamizo-Gonzalez et al., (2015) stated that mining educational data can offer HEI valuable insights that can inform strategic decision-making regarding resource allocation for educational excellence.

In the context of online learning, most of what is known about student engagement usually comes from self-reported student data but analyzing and mining LMS data has the potential to provide a new perspective on their online engagement. Therefore, to gain insight into and understanding of the current usage of LMS and engagement within LMS systems, the goals of this research are to apply EDM techniques to evaluate LMS usage to discover users' behavior usage and engagement patterns, as well as to identify some indicators that may help in measuring the level of engagement. As many previous studies have acknowledged, these usage patterns and engagement behaviors can help clarify how instructors and students engage with certain learning tools and activities, allowing corrections or improvements to be made for aspects that are not being used properly (Dahleez et al., 2021; Lee et al., 2019; Viloria Silva et al., 2019).

Furthermore, identifying other engagement factors supported by the literature review as well as studying their level of influence on the students' online engagement would provide a full understanding of all aspects that stand behind the low level of engagement in the online environment (Saa, Al-Emran, & Shaalan, 2019; Shah & Cheng, 2019). Therefore, this research will also apply another method called DEMATEL technique to model the causal relationship between the identified factors based on the perceptions of the key stakeholders (instructors and students). DEMATEL technique is an effective approach that collects relevant knowledge, analyzes the interrelationships among factors in complex systems or domains, and visualizes this structure through a cause–effect relationship diagram (Adegoke et al., 2021; Roostaie & Nawari, 2022). With this understanding, this can better improve online learning and activities that lead to changes in student engagement level in such environments.

1.2 Background of the Study

In this information age, one of the most influential HE institutions is universities (Manek, Vijay, & Kamthania, 2016). Over the past decade, evaluating universities was based on indicators such as teaching, size, GPA average, and annual alumni numbers. Although this type of information is important, it is also crucial for these universities to collect and understand information related to the use of educational online systems offered to facilitate learning processes such as LMS and how students engage with courses in this environment (Usher et al., 2021). The literature indicated that student engagement has become synonymous with measuring the quality of learning and teaching in universities (Lee et al., 2019). Hence, currently, one of the major challenges faced by the universities is how to improve and manage the learning processes to be more effective through the interpretation and analysis of big data stored in their educational systems (Duangekanong & Huang, 2022; Mense et al., 2020). Big data emphasizes that the data itself is a way to value any university learning system and it is also an important value for HEI (Berwind et al., 2016; García & Secades, 2013). To reach this goal, EDM is currently one of the most appropriate approaches in providing clear insights and a better understanding of the reasons for low engagement within LMS systems that can be gained by grouping LMS courses according to the students' level of engagement in online activities. This can help university decision makers such as faculty management and instructors to make better decisions about their teaching activities based on the information provided.

LMS is one of the important spaces where EDM can be applied to discover useful information that is valuable in mapping student engagement. Most LMS systems contain different types of tools that allow students to participate in online course activities and leave traces that can be used to analyze their communications and interactions. Each action is part of a response pattern of students who posted more messages in discussion forums, who submitted more assignments or quizzes, what kind of resources they accessed more often, and so on. These different actions (activities or behaviors) in the online environment leave a data track of where they have been and when. The analysis of the data provided by the LMS is essential for improving instructional methods and for developing new teaching strategies or modifying the methodologies used (Al-Nuaimi et al., 2022; Hamid et al., 2022). By mining this data, the tacit structure of the system can be visualized using any open-source tools. The same procedure could be applied to analyzing instructors' data to understand how students engage with the instructors and vice versa. Those with more postings are more engaged and use the systems frequently. In this vein, the first objective of this research aims to apply EDM techniques to find useful and hidden information and acquire knowledge from existing LMS data, identify different measures to guide improvements in the learning process, and to support university administration to better understand how stakeholders interact and engage with the courses and activities within these systems. The categorization can be carried out automatically using statistical mining methods and clustering algorithms.

However, it is important to inform that engagement is a theoretical concept and there are many other factors that cannot be understood from LMS data only. Therefore, a systematic review is also conducted to examine the other factors influencing student engagement in online learning to support the indicators obtained from the factual data of LMS. Also, there is still a need to understand the most critical factors that cause low levels of student' online engagement and only factors with strong relationships should be taken into consideration. The use of DEMATEL approach in modelling the key factors affecting students' engagement in online environments aims at providing a ''strategic mapping'' to better understand the cause-and-effect relationships among these factors (Adegoke et al., 2021). This can also help in providing an in-depth understanding of the main factors that cause low level of online engagement (Fredricks et al., 2016).

1.3 Problem Statement

Despite the wide implementation of LMS, instructors' and students' utilization of LMS is still minimal and this utilization is still not within its full potential even after Covid-19 pandemic (Seraji et al., 2022). Several prior studies reported that students' and instructors' use of the LMS is very rare and mainly for delivering or downloading learning resources (Al-Nuaimi et al., 2022; Al-Sharhan et al., 2020; Bradley, 2021; Ghilay, 2019). Accordingly, a low usage level of LMS indicates a low level of engagement with activities within the system. Therefore, instructors' and students' engagement are also seen as relatively low within these systems (Avcı & Ergün, 2022; Christopoulos et al., 2018; Ekanayake & Weerasinghe, 2020). Thus, exploring LMS data to obtain more information about their usage behaviors and engagement level is critical to improving online learning.

During the current decade, the number of online courses is increasing, and understanding how students engage and interact with these courses and related activities in the online environment has become an ongoing problem in universities (Maslov et al., 2021; Sun et al., 2022). To help in solving such problem, many researchers have held on trying to study student engagement in online environments from different perspectives and contexts to increase the advantages of online learning, and many of these studies have found that limited engagement in online learning systems to be a persistent and widespread problem (Duangekanong & Huang, 2022; Jain et al., 2013; Lee et al., 2019; Shah & Cheng, 2019; Sun et al., 2022; Wells et al., 2016). Furthermore, measuring students' engagement in the online environment using traditional surveys and qualitative methods such as Student Engagement Questionnaire (SEQ), National Survey of Student online Engagement (NSSE), as well as the College Student Experiences Questionnaire (CSEQ) is difficult to generalize or scalable, interrupt students, require a significant amount of time, and they are often not a good option for measuring student engagement in the online environment. Meanwhile, the measurement of students' online engagement based on the data generated from the online systems that can provide complete insights and statistics based on the real data for overall evaluation has not been well researched (Hussain et al., 2018; Nkomo & Nat, 2021). Therefore, many recent studies are still constantly calling for further studies that are able to deeply analyze and evaluate the LMS usage to discover hidden behavior patterns especially regarding the problematic nature of students' engagement in the online environments (Avcı & Ergün, 2022; Dahleez et al., 2021; Ismail et al., 2021; Lee et al., 2019; Teng & Wang, 2021). These prior studies mainly emphasized on the need for new intelligence methods that could help in this matter.

Although LMS provides some reporting tools that provide very limited data reporting options to produce statistical reports, these tools are not very specific that help to draw useful conclusions either for the course activities, users interaction, or for decision makers (Rajabalee & Santally, 2021). In addition, no guidance is available for decision makers to understand any of the many available data points indicate students' engagement in educationally purposeful activity that may contribute to enhance online learning (Mense et al., 2020; Seraji et al., 2022). Meanwhile, the amount of online courses data stored in educational databases is rapidly increasing which is of great value for analyzing students' and instructors' usage and engagement based on every mouse click within the system. This means that the fundamental measure of users' engagement within LMS is the degree to which they interact and use the system. The ability to manipulate and utilize such data to discover useful patterns is considered a significant strategy for the future plan of online learning (Mthethwa-Kunene & Maphosa, 2020; Martin & Bolliger, 2018). Accordingly, in the last few years, researchers have begun investigating various techniques of EDM for analyzing LMS data and extracting information and knowledge to support decision-making to improve educational systems (Hamid et al., 2022; Ismail et al., 2021; Martinez-Abad & Chaparro-Caso-López, 2017; Moubayed et al., 2020; Varela et al., 2019).

Although the advance in the field of EDM makes it possible to mine LMS data to improve the quality of online learning (Hernández-Blanco et al., 2019; Lee et al., 2019), there is a dearth of research on the application of EDM in university environment to discover hidden patterns or aspects that are difficult or impossible to capture by other means (Mense et al., 2020). Furthermore, the focus of existing EDM studies was mostly on predicting students' performance and achievement or predicting early dropout and not many studies have been conducted on student engagement (Asif et al., 2017; Casey & Azcona, 2017; Márquez-Vera et al., 2016; Saa, 2016; Varela et al., 2019). Therefore, to contribute to solving problems associated with students' online engagement, the first objective of this study is to apply EDM techniques that can help in understanding LMS usage behavior patterns of students and instructors and their engagement level in the online environment. This requires exploring their online activities (statistically and visually) based on the log files in LMS to find any useful information that can be used for further improvement. Furthermore, this also requires categorizing (clustering) the various online courses according to students' interaction in the online activities which may show different patterns that help in estimating the level of online engagement in such courses. Understanding and categorizing such data reveal common characteristics, indicators, usage patterns, and behaviors among students that university decision makers may be most attracted to.

However, although LMS data describing students' activities linked with their academic learning that can provide useful indicators of student engagement in the online environment, this is not sufficient to obtain a full understating of the issues surrounding students' online engagement (Saa et al., 2019; Shah & Cheng, 2019). Based on the literature, there are many other influencing factors beyond the behavior patterns such as personal factors (e.g., prior experience, academic skills and abilities), academic, and environmental factors that vary from one study to another and there is no agreement on which of these factors has a greater impact on student engagement in online environments. The literature indicated that these problems have been the focus of many researchers over the past few years (Christopoulos et al., 2018; Elshami et al., 2022; Kordrostami & Seitz, 2022; Lee et al., 2019; Shah & Cheng, 2019). Hence, to fill this methodological gap in the current literature, this research also aims to go some way in responding to the growing call from online learning scholars to explore the key factors affecting students' online engagement through the literature review and experts' opinions and model the causal relationships among these factors. Studying such factors is critical to providing a full description of all the factors that surround students' online engagement. Mainly, to answer two research questions: 'What are the key factors affecting students' engagement in online courses?' and 'What are the causal relationships between these factors?'. To answer these two questions, DEMATEL technique was used to model the causal relationships between the factors and to build an impact-relation map from the students' and instructors' perspectives.

1.4 Research Objectives

This study aims to conduct an exploratory analysis of a dataset retrieved from the LMS system to identify usage patterns, engagement indicators, and user behaviors to provide insights to the decision makers about the current usage and factors linked to students' online engagement. In addition, this study also aims to examine the literature to collect the common factors affecting student online engagement to study the causal relationships between the factors and identify the key factors. Hence, the main objectives of this study are summarized as follows:

1- to evaluate the usage behavior and engagement of students and instructors in the LMS system using EDM techniques (statistical and visual data mining).
2- to identify the significant characteristics and indicators of courses with different levels of online engagement using the clustering technique.
3- to model the causal relationships between preidentified factors using DEMATEL technique from the students' and instructors' perceptions.
4- to identify the key factors affecting students' engagement in the online courses.

1.5 Research Questions

Based on the research problem and study objectives, the research questions are as follows:

1- What novel and useful knowledge can be discovered from the usage behavior and engagement patterns of students and instructors in the LMS systems?
2- What are the significant characteristics and indicators of the courses with different levels of online engagement based on clustering technique?
3- What are the causal relationships between preidentified factors from the students' and the instructors' perceptions?

4- What are the key factors affecting students’ engagement in the online environment?

1.6 Research Framework of the Study

This research is grounded on three main theories (concepts) involving engagement, EDM, and DEMATEL theory. This section is to provide the theoretical background of these three main concepts. However, as the engagement concept, theories, and models will be discussed in detail in Chapter 2, the focus here is on the theoretical concept of EDM and DEMATEL method. In addition, as EDM is based on the theory of Data Mining (DM), a brief explanation is provided first about DM process and the relevant terms and models.

1.6.1 Data Mining Theory

DM or knowledge discovery in databases (KDD) is defined as the process of extracting hidden predictive information and discovering interesting patterns and knowledge from large databases and summarizing it into useful information (Dutt et al., 2017). Precisely, DM is one of the five stages of the KDD process and it mainly relates to the means by which the patterns are extracted and interpreted from data (Fayyad et al., 1996). The five stages of KDD involving Selection (creating a target data set, variables, data samples on which discovery is to be made), Pre-processing (data cleaning and pre-processing), Transformation (transformation of the data using transformation methods), Data Mining (searching for patterns of interest based on the data mining goal), and finally Interpretation/Evaluation of the mined patterns. Figure 1.1 illustrated the five stages of KDD process.

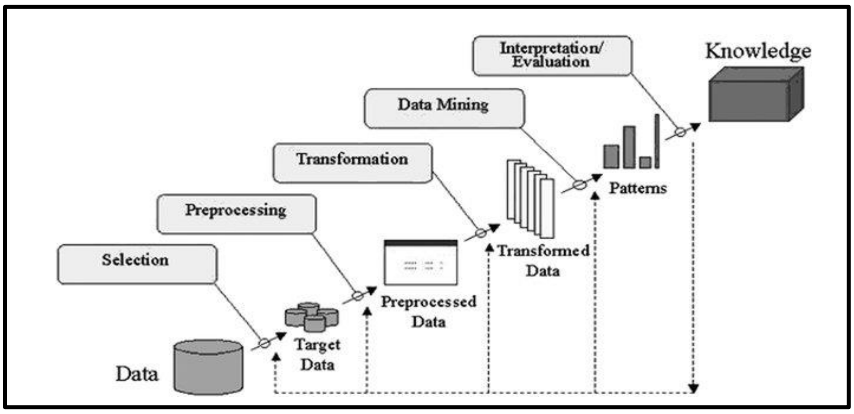

Figure 1.1 The five stages of KDD (Fayyad et al., 1996)

1.6.2 Educational data mining (EDM)

EDM is the application of DM techniques, tools, and algorithms in educational context, which is defined as the development of methods and techniques to discover knowledge and the unique types of data generated in educational settings. EDM stands up from several related statistical and computational approaches that form a set of data mining methods. There are many standard models and methodologies to implement the analysis in DM/EDM such as SEMMA method (Milley & James, 1998), Data Mining Model (Srivastava et al., 2000), and CRISP-DM reference model (Shearer, 2000).

These methods and models focus on the technical aspects rather than the relationship to the task domain and the process itself (Peng et al., 2008). However, to structure these methods, the research and practice in DM have extended to include steps along the DM model (Srivastava et al., 2000). According to the DM model shown in Figure 1.2, mining web access logs is a process that consists of three sequential steps: data gathering, pre-processing, and pattern analysis. Data gathering and pre-processing steps to filter and format the log entries as well as pattern discovery that consists of using a variety of techniques and algorithms such as classification, clustering, association rule mining, and sequential pattern analysis on the transformed data to find useful patterns. Meanwhile, pattern analysis is a process in which the discovered patterns are interpreted and retrieved (Murnion & Helfert, 2011; Zaiane & Luo, 2001). This model is described as the technical perspective of DM and was commonly used in EDM research (García et al., 2007; Zaiane & Luo, 2001).

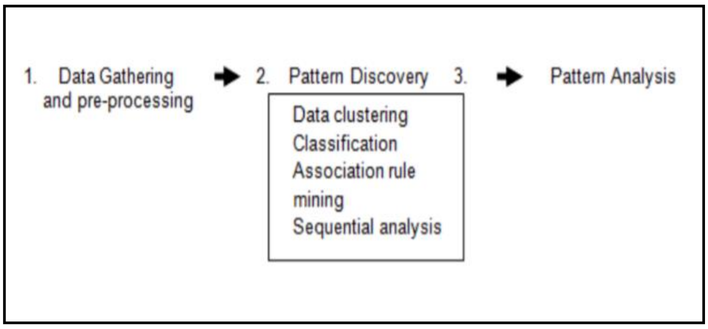

Figure 1.2 Data Mining Model (Srivastava et al., 2000)

SEMMA is another method that can be used as a reference model to guide the user on the applications of the EDM process. It was developed by the Statistical Analysis Systems Institute (SAS) and consists of a list of steps in a five- stages cycle. SEMMA is the acronym that stands on Sample (data sampling), Explore (data exploration by searching for unexpected anomalies and trends to gain understanding and insight), Modify (data modification by creating, selecting, and transforming the variables), Model (data modeling by allowing the software to automatically search for a set of data that reliably predicts the desired outcomes), and Assess (assessing the data by evaluating the reliability and utility of the results from the DM process and estimating how well they are performing). The five stages of SEMMA are shown in Figure 1.3.

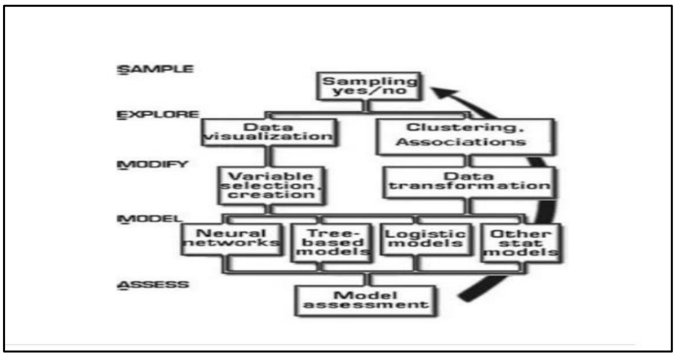

Figure 1.3 The five stages of SEMMA (Milley & James, 1998)

However, as DM matured as a discipline and its applications were applied to a wide range of problem domains, these technical steps were incorporated into a more comprehensive model developed by Shearer (2000) called the CRISP-DM reference model. CRISP-DM model which stands for CRoss Industry Standard Process for Data Mining is also one of the best-known models that defines the tasks and processes to implement successful DM projects. This model consists of a sequence of steps that are also known as the data mining cycle.

Both SEMMA and CRISP-DM can be used as reference models for the implementation of DM methods. However, the CRISP-DM model is more complete than SEMMA and was the most preferred methodology in this field (Azevedo & Santos, 2008). Therefore, it was chosen as a reference model in this study due to its widespread use by researchers in the educational field over the past ten years (Kabakchieva, 2013; Murnion & Helfert, 2011). The CRISP-DM comprises six phases as shown in Figure 1.4. The sequence of the phases is not rigid. In addition, DM does not end once a solution is deployed. Moving back and forth between different phases is always required. The lessons learned during the process and from the deployed solution can pose new and more-focused questions. Subsequent DM processes will benefit from the previous operational experiences (Hofmann & Tierney, 2009). As shown, the outcome of each phase determines which phase or task of a phase should be implemented next.

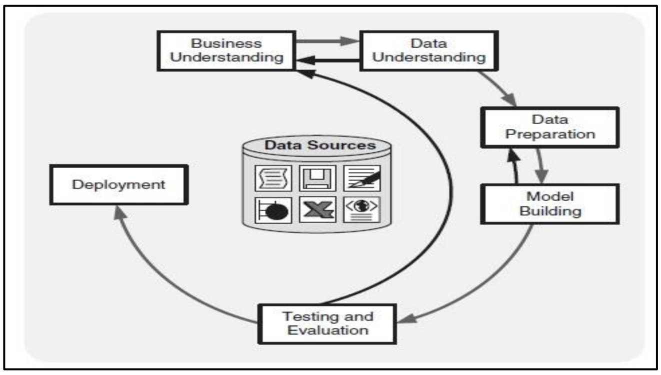

Figure 1.4 CRISP-DM reference model (Shearer, 2000)

As demonstrated, the first phase, Problem definition, is the start of the cycle. Based on the understanding of the domain problem, a DM problem or hypothesis can be derived. The next four phases from data exploration to evaluation establish the technical steps for data mining , resulting in a model, knowledge or information which can be deployed. Finally, in the Deployment phase, the results of the data mining are used to solve the originally defined problem. In order to fully implement the cycle, the deployment and problem definition phases must be appropriately linked, as the deployment results must be evaluated and returned to the problem definition for the next iteration of the cycle (Chapman et al., 2019; Murnion & Helfert, 2011). The six phases of CRISP-DM are described in detail in Chapter 3.

1.6.3 DEMATEL Theory

The theory of the DEMATEL (Decision Making Trial and Evaluation Laboratory) method is built on the basis of exclusively directed graph theory known as causal map diagrams or matrixes. This graph is more effective as compared to undirected graphs because it is able to represent directed links of the subsystems or problems (Liu, Guo, & Zhang, 2019). It was originally developed by the Science and Human Affairs Program of the Battelle Memorial Institute of Geneva between 1972 and 1976 to solve difficult and complex problems (Fontela & Gabus, 1976; Gabus & Fontela, 1973). It can transform the interrelations between factors into an understandable visual structural model of the system and divide them into two groups called a cause group and an effect group. The main goal of constructing a DEMATEL model for any problem is to be able to predict the outcome by allowing related factors or issues to interact with each other.

Therefore, to date, DEMATEL technique was widely used as one of the most operative and efficient tools for solving the cause–effect relationships between causally correlated evaluation factors or variables. It is mainly applicable and useful for analyzing and characterizing the interdependent relationships between factors in a complex system and classifying them for making long-term strategic decisions and identifying areas for improvement (Maqbool & Khan, 2020; Roostaie & Nawari, 2022; Si et al., 2018). Besides, this technique can help decision-makers understand the hidden causal relationships that could contribute to more relevant and important solutions. Finally, this technique can help decision makers improve their decision capabilities and enhance strategic thinking (Si et al., 2018).

The graphical illustration of DEMATEL is a set of nodes connected by edges. Each node represents a factor (F) related to a particular domain to describe the behavior of the system or problem. The arrows represent the causal relationships (impacts) between factors. The causal effect between each pair of factors can be mapped out by drawing the causal map diagram. For example, an arrow from F1 to F3 demonstrates the influence that F1 has on F3, and the strength of its effect is 1. More explanations about the DEMATEL technique and how it works are discussed in Chapter 3. Figure 1.5 represents a simple graphic illustration of DEMATEL technique.

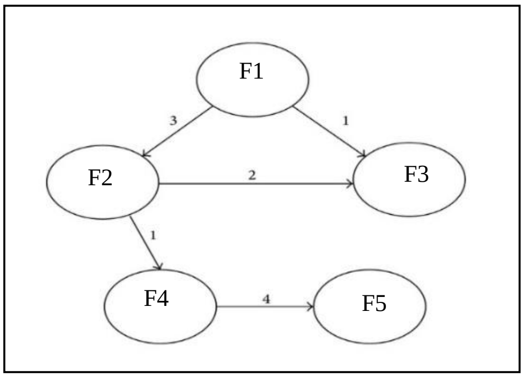

Figure 1.5 A simple illustration of the DEMATEL causal map

The findings from the two methods used in this study (EDM and DEMATEL technique) allow a full description of all factors and issues surrounding students' engagement in the online environment. On this basis, the framework of the study is shown in Figure 1.6.

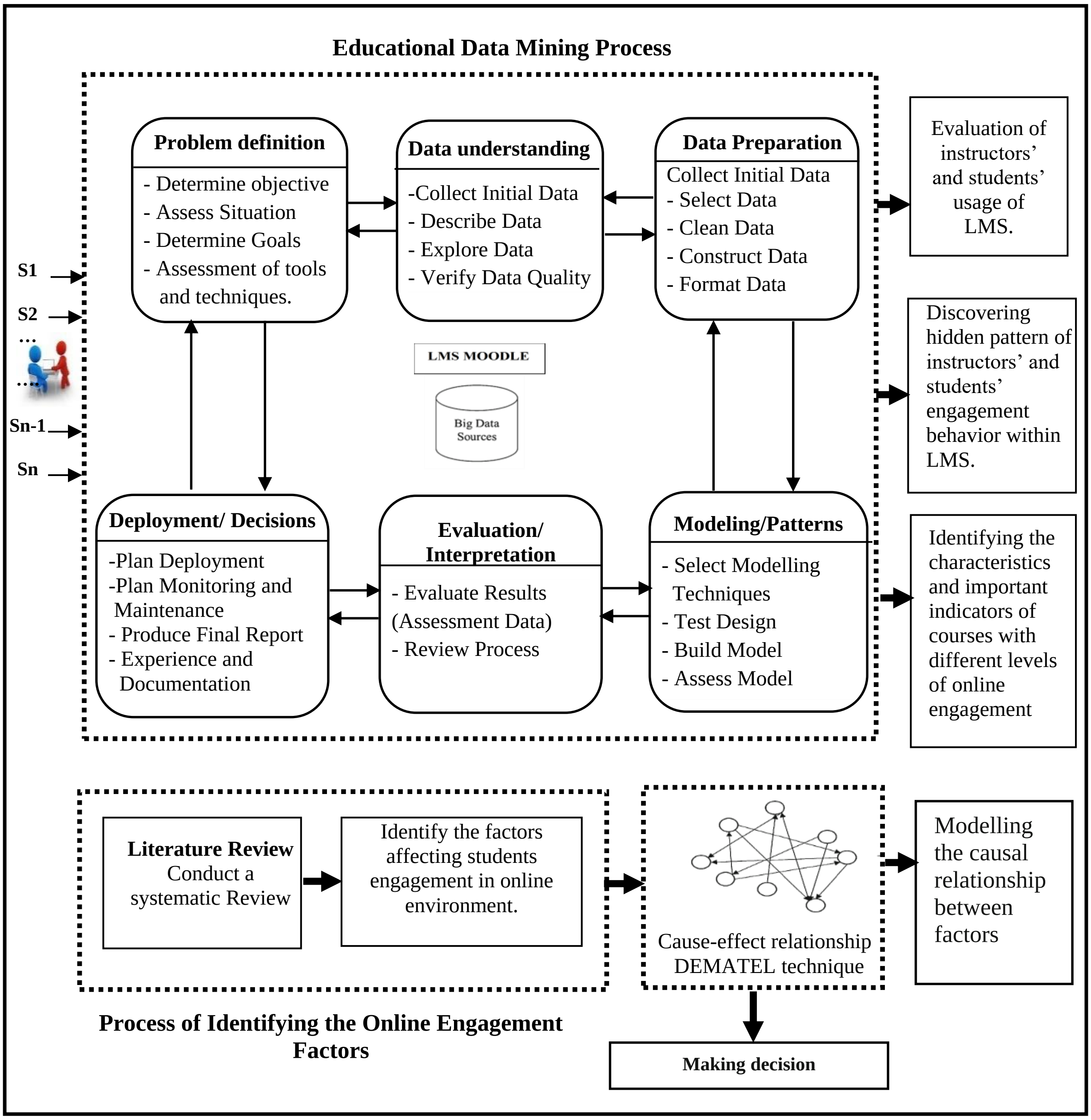

Figure 1.6 Research framework of the study

1.7 Significance of the Study

The emerging application of EDM is becoming a new useful approach for a new era. Accordingly, EDM and big data in educational systems have a significant role to play in the future of higher education. For this, the significance of this study is twofold. First, the analysis and evaluation process of instructors' and students' usage behaviors and engagement patterns within LMS using EDM help decision makers to gain insights into the online learning process and discover useful patterns about the level of engagement for further improvement (Al-Fraihat et al., 2020; Zanjani et al., 2016). Evaluating LMS usage and engagement is also an important process for universities to improve their courses according to the provided indicators (Campagni et al., 2015).

Investigating student engagement in an online learning environment based on LMS data and using EDM techniques as they have the ability to change the existing conceptions and the measurement of engagement will enable universities to understand how the different elements or factors of online learning experience affect student engagement (Hamid et al., 2022). With this understanding, universities can better improve online courses, learning strategies, and qualities that lead to transformation in student engagement in the online environment. In addition, along with improving learning and teaching mechanisms, the ability to successfully evaluate students' online engagement is essential as a key towards truly improving the online learning and helping universities maintain a positive educational atmosphere, and develop innovative plans that take into account student online engagement based on the new insights presented (Duangekanong & Huang, 2022). Moreover, it can help motivate instructors to use LMS regularly that will positively reflect on students' use as well (Kordrostami & Seitz, 2022). These new insights can also help university decision makers to derive appropriate treatment and policies to the instructors and students to make more effective information-based decisions that ultimately support the educational environment for positive outcomes (Moubayed et al., 2020).

Secondly, understanding the operational elements or core factors that affect student engagement will be important for the future design of an effective online environment. Therefore, this study will significantly contribute to filling this methodological gap by employing smart methods such as DEMATEL technique to provide an in-depth understanding of the key factors that play an important role on students' level of engagement in online courses. This will also help decision makers to understand the core factors and the causal relationships responsible for low level of engagement in online courses and significantly affect the efficiency of LMS systems as well. By studying the cause-and-effect relationships between these factors, only factors with strong relationships will be taken into consideration for further improvements. This can help in giving a clear insight for educational policy makers and system designers to apply the necessary interventions and measures to rectify the situation. This combination of data-processing using EDM and mapping technique using DEMATEL is an aid to significantly improve online learning in higher education and defining the path to be followed in the new educational era.

1.8 Operational Definitions

1.8.1 Learning Management System (LMS)

LMS is a comprehensive web-based platform designed for monitoring, documentation, reporting, delivery of courses, tracking, assessment, and reporting of student progress, user management, online courses administration in traditional classrooms, e-learning, or a combination of the two (blended learning), and other services for educational institutions (Ghilay, 2017). LMS also provides an environment and place for student-student, student-instructor, and student-content interactions to occur and tracking and assessment of student activities. These systems collect each student and instructor online behaviors data in every course. One of the key reasons for choosing the LMS for this research is its ability to collect and analyze detailed data on student and instructor online behaviors. The system tracks and records each student and instructor's actions and movements within the LMS, providing valuable insights into their engagement patterns and usage behaviors. This data can be analyzed and interpreted to understand students' interactions with different course activities and LMS tools over a two-year study period. In this study, specifically, ***eLearn@usm.my*** is the chosen LMS because it serves as the official e-learning environment at the Universiti Sains Malaysia (USM). This choice is based on several factors. Firstly, ***eLearn@usm.my*** offers comprehensive features that cater to the diverse needs of educational institutions. Secondly, the LMS has the capability to track and analyze user behaviors, enabling the study to examine the frequency of actual LMS usage for educational purposes and track patterns of behavior. Finally, the adoption of ***eLearn@usm.my*** as the official e-learning environment at USM ensures its widespread use and familiarity among students and instructors. By leveraging ***eLearn@usm.my***, this research aims to gain insights into students' interest and desire in specific content or features within the LMS and evaluate the effective utilization of these features.

1.8.2 Online environment

Online environment refers to the digital platform or virtual space in which learning activities, interactions, and resources are accessed and facilitated through internet-based technologies, such as LMS or online educational platforms. It encompasses the digital ecosystem where online courses, collaborative tools, instructional materials, and communication channels are utilized for educational purposes.

1.8.3 LMS Usage behaviors, Engagement behaviors, and Engagement Level

LMS usage behaviors and engagement patterns in the context of this study are defined as users' actions and movements that students and instructors make when they use or navigate the system and interact with different course activities and LMS tools. It is also referred to the frequency actual use of the LMS system by both users for educational purposes. Tracking patterns of behavior and measuring interaction on the LMS system enable researchers to determine the users' interest and desire in specific content, or certain features and whether those features are being properly exploited. A brief explanation of the differences between the three terms are as follows:

1.8.3.1 Usage behaviors

Usage behavior refers to the actions and patterns of system usage by users within the LMS. It includes tracking and analyzing factors such as frequency of logins, tools or features accessed, and overall usage patterns. It focuses on understanding how users interact with the LMS in terms of their actions and usage patterns.

1.8.3.2 Engagement behaviors

Engagement behavior encompasses the involvement, interaction, and interest demonstrated by users within the LMS. It goes beyond usage and focuses on the key aspects of user engagement. This includes factors such as active participation in discussions, completion of assignments, collaborative activities, and interactions with course content and peers. Engagement behavior provides insights into the depth and quality of user engagement within the learning environment.

1.8.3.3 Engagement Level

In this study, the level of engagement is operationally defined as the collective measure of user involvement and interest exhibited within the LMS. It encompasses both usage behavior and engagement behavior, providing a holistic view of how actively and meaningfully users interact with the course activities and overall learning experience. The level of engagement is assessed through the total number of clicks (log data records) that signify the extent of interaction between students and instructors within the LMS. This comprehensive definition ensures a thorough description of students' and instructors' levels of engagement, considering the classification and distinctive characteristics of each group. It encompasses a diverse range of interactions, including logs, posts, views, and other forms of engagement with online learning resources, course activities, individuals, and various LMS tools. By incorporating these various types of engagement, a comprehensive understanding of participants' involvement and interaction within the LMS environment is achieved, facilitating a nuanced analysis of their levels of engagement.

To further understand the level of engagement, it will be classified into different categories (clusters) based on the detectable characteristics of participants' behaviors and the number of activities. Each category reflects a distinct level of involvement and interaction within the LMS. The identification of the level of engagement is based on the characteristics of each cluster, and the number of clusters is determined automatically using the results of the clustering technique and evaluation measurements. By analyzing the unique features of each cluster, insights into the varying levels of engagement exhibited by participants are obtained. For instance, participants showing very high level of engagement demonstrate high levels of interaction, extensive usage of LMS tools, and frequent contributions to discussions and collaborative activities. On the other hand, those classified as low level of engagement show minimal interaction, limited usage of LMS tools, and infrequent participation in course activities. This approach enables a comprehensive understanding of engagement patterns and provides a foundation for targeted interventions and enhancements.

1.8.4 Online Engagement

Online engagement refers to students' and instructors' interaction, participation, and involvement in online learning activities in the LMS (Millar et al., 2021). While students' online engagement is defined as their active participation in online course activities, interaction with peers, instructors, and content, as well as collaboration with others to construct knowledge using different LMS tools to achieve learning goals, instructor engagement is defined as the instructor's role, presence, and active participation in online courses, interaction with students and content, and providing prompt feedback to help guide students during course activities using LMS tools to achieve learning goals.

1.8.5 Engagement Factors

Engagement factors in the context of this study refer to characteristics that generate a high level of interest and motivation in the learning process. The presence or absence of these factors can impact students' engagement negatively (Shah & Cheng, 2019). In online courses, engagement factors encompass interactions with the LMS and other users, leading to increased use, interaction, engagement, and acceptance of the environment, ultimately enhancing learning outcomes (Groves et al., 2015). Student engagement is influenced by resource allocation, learning strategies, and services provided by universities to encourage active participation in learning activities. Furthermore, key factors affecting student engagement in the online environment are those with significant influences, while secondary factors are directly or indirectly affected by the key factors and should be considered accordingly.

1.8.6 Educational Data Mining

Educational Data Mining (EDM) is the process of extracting valuable insights and patterns from educational data using computational techniques to enhance teaching and learning practices. In the context of this study, EDM was utilized to analyze and interpret large-scale educational data related to students' and instructors' activities and engagement in the LMS for providing valuable information for improving instructional practices and decision-making. This allowed for a deeper understanding of student learning behaviors, instructional effectiveness, and educational outcomes in the online setting.

1.8.7 DEMATEL Technique

The DEMATEL technique (Decision-Making Trial and Evaluation Laboratory) is a quantitative method used to analyze and assess the causal relationships and interdependencies among factors or variables. In the context of this research on identifying factors affecting student engagement in the online environment, the DEMATEL Technique is employed to determine the relative influence and impact of various factors on student engagement.

1.9 Summary

This chapter presented the main elements of the research that include the background of the study, problem statement, research objectives, research questions, conceptual framework, and significance of the study. The operational terms that were used throughout the research were also defined.