Detecting Students At-Risk Using Learning Analytics

A thesis
submitted in fulfilment
of the requirements for the Degree
of
Doctor of Philosophy

By
Ayman Saleh S Albassam

Supervised by
Professor Katrina Falkner
Associate Professor Claudia Szabo

School of Computer Science
University of Adelaide
June 2019

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Ayman Saleh S Albassam

This thesis is dedicated to my beloved parents, my lovely wife and little son

Acknowledgements

First of all, I would like to express my gratitude to my supervisors, Professor Katrina Falkner and Associate Professor Claudia Szabo for their support, guidance, and feedback throughout the period of this research. Also, I am immensely grateful to the members of the Computer Science Education Research Group for their invaluable support during my studies. I would like also to thank my close friends, HDR colleagues and university staff who create a great research atmosphere and have helped me throughout my PhD journey.

I also express my sincere appreciation to Umm Al-Qura University, Kingdom of Saudi Arabia, who supported me financially during my PhD journey.

I am grateful to Alison-Jane Hunter for her professional support in editing this thesis.

Last, but not least I would like to express my heartfelt gratitude to my beloved parents, sisters and brothers, lovely wife and close friend, Abdullah Alsalomi, who never stopped encouraging and supporting me. This thesis was not possible without you.

Abstract

The issue of supporting struggling tertiary students has been a long-standing concern in academia. Universities are increasingly devoting resources to supporting underperforming students, to enhance each student,s ability to achieve better academic performance, alongside boosting retention rates. However, identifying such students represents a heavy workload for educators, given the significant increases in tertiary student numbers over the past decade.

Utilising the power of learning analytic approaches can help to address this problem by analysing diverse students' characteristics in order to identify underperforming students. Automated, early detection of students who are at potential risk of failing or dropping out of academic courses enhances the lecturers' capacity to supply timely and proactive interventions with minimal effort, and thereby ultimately improve university outcomes.

This thesis focuses on the early detection of struggling students in blended learning settings, based on their online learning activities. Online learning data were used to extract a wide range of online learning characteristics using diverse quantitative, social and qualitative analysis approaches, including developing an automated mechanism to weight sentiments expressed in post messages, using combinations of adverbs, strengths. The extracted variables are used to predict academic performance in timely manner.

The particular interest of this thesis is on providing accurate, early predictions of students, academic risk. Hence, we proposed a novel Grey Zone design to enhance the quality of binary predictive instruments, where the experimental results illustrate its positive overall impact on the predictive models, performances. The experimental results indicate that utilising the Grey Zone design improves prediction-accuracy by up to 25 percent when compared with other commonly-used prediction strategies.

Furthermore, this thesis involves developing an exemplar multi-course early warning framework for academically at-risk students on a weekly basis. The predictive framework relies on online learning characteristics to detect struggling students, from which was developed the Grey Zone design. In addition, the multi-course framework was evaluated using a set of unseen datasets drawn from four diverse courses (N = 319) to determine its performance in a real-life situation, alongside identifying the optimal time to start the student interventions. The experimental results show the framework,s ability to provide early, quality predictions, where it achieved over 0.92 AUC points across most of the evaluated courses. The framework's predictivity analysis indicates that week 3 is the optimal week to establish support interventions.

Moreover, within this thesis, an adaptive framework and algorithms were developed to allow the underlying predictive instrument to cope with any changes that may occur due to dynamic changes in the prediction concept. The adaptive framework and algorithms are designed to be applied with a predictive instrument developed for the multi-course framework. The developed adaptive strategy was evaluated over two adaptive scenarios, with and without utilising a forgetting mechanism for historical instances. The results show the ability of the proposed adaptive strategy to enhance the performance of updated predictive instruments when compared with the performance of an unupdated, static baseline model. Utilising a forgetting mechanism for historical data instances led the system to achieve significantly faster and better adaptation outcomes.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

For many years, the problem of improving students' academic performance has been a consistent concern in higher educational contexts (Arnold & Pistilli 2012, Astin 1984, Burgos et al. 2018, Clark & Ramsay 1990, Jayaprakash et al. 2014, Johnston 1997, Pantages & Creedon 1978, Spady 1970, Tinto 1975, Wilson 2005, Wong 2017). A variety of approaches have been used to enhance the students' academic achievements, including identifying and supporting struggling students in a timely and effective manner (Wong 2017). This raises challenges related to monitoring learning progress and identifying poorly performing students in a timely fashion in order to deliver practical academic support. Early detection of potentially underperforming students helps instructors to intervene effectively to address their learning challenges. Providing timely and meaningful academic support to struggling students can lead to improvements in the quality of graduates' achievements, as a result of the improvements in students' academic performance. Previous researchers have confirmed that providing early academic interventions to students in need is reflected in the quality of graduates, in terms of increases in retention rates (Arnold & Pistilli 2012, Burgos et al. 2018, Smith et al. 2012) and positive overall learning outcomes (Cassells 2018, Dodge et al. 2015, Jayaprakash et al. 2014).

In higher education settings, numerous studies have examined the effects of providing early academic support to students who have performed poorly in their courses at an institutional level. Sclater et al. (2016) report several international case studies conducted in Australia, the UK and the US to evaluate the effectiveness of diverse academic interventions on academic outcomes. These case studies demonstrate the positive impact of early interventions on students' academic achievements. Moreover, in a comprehensive review of the impact of academic interventions on students' success, Wong (2017) stated that providing academic interventions improved students' outcomes over non-intervention groups of students in all the higher education institutions reviewed, however, the effect size of the interventions was varied across the institutions.

Although the majority of studies in the literature recorded the significant impact of early academic interventions on students' outcomes, others logged limited effects. For instance, in the empirical studies conducted by Arnold & Pistilli (2012) and Smith et al. (2012), they found significant positive impacts from applying early interventions on student retention rates. Furthermore, Cassells (2018) and Jayaprakash et al. (2014) studied the contribution of timely interventions on groups of students who had been labelled as at academic risk. Their studies indicate that those students who were involved in the study achieved higher grades over nonintervention sets of students by eight and six percent respectively. However, other studies have recorded the limited influence of early academic support on students' academic performance (Dawson et al. 2017, Dodge et al. 2015).

The interest in identifying factors associated with academic risk has led many researchers to investigate these phenomena for over six decades (Pantages & Creedon 1978). The higher education literature widely recognises the term 'at-risk students' as those who face a high risk of underperforming academically and eventually fail or withdraw from academic courses (Azcona & Casey 2015, Cassells 2018, Falkner & Falkner 2012, Jayaprakash et al. 2014, Wolff et al. 2013). However, several previous studies utilised one aspect of the definition by referring to academic risk as either at attrition risk alone (Agnihotri & Ott 2014, Chai & Gibson 2015, He et al. 2015), or at being at-risk of performing poorly in academic courses (Bainbridge et al. 2015, Choi et al. 2018, Dodge et al. 2015, Rogers et al. 2014, Wang & Newlin 2002). For this thesis, 'at-risk students' are defined as those students who have both a high potential for failure and/or may withdraw from their academic courses.

Early attempts identified several factors associated with students' academic risk characteristics (Astin 1984, Baker & Siryk 1984, Clark & Ramsay 1990, Everett & Robins 1991, Gerdes & B 1994, McKenzie & Schweitzer 2001, Pantages & Creedon 1978, Spady 1970, Terenzini & Pascarella 1978, Tinto 1975). At this time, the focus was on demographic characteristics, pre-enrolment performance and other psychosocial factors to identify university outcomes. While the factors studied can provide a general idea about students' overall learning characteristics and academic performance, these factors have very limited ability to indicate an individual student's actual learning progress in a particular course.

Therefore, other researchers have considered utilising course-related elements. Data drawn from course-specific learning activities can specify how well students engaged in undertaking a particular course. For instance, researchers employ continuous assessment attributes as indicators of students' performance in course contexts (Mayilvaganan & Kalpanadevi 2014, Minaei-Bidgoli et al. 2003). Although utilising traditional in-between assessments data allows educators to identify and support students who are at-risk in a particular subject, assessment data frequently becomes available too late, minimising the opportunities to provide proactive interventions (Almosallam & Ouertani 2014, Macfadyen & Dawson 2010).

On the other hand, the integration of advanced technologies within traditional learning processes has provided a new source of data that may expose previously hidden aspects of students' learning patterns and consequently help us achieve better understanding of the factors affecting academic performance. Thus, higher education organisations now utilise digital learning footprints, where tremendous volumes of detailed data are recorded every day about students' learning activities, in order to recognise students' learning progress whilst they are undertaking courses.

The availability of students' online learning data impacts positively on educators' capacity to discover learning patterns, enabling them to track learning progress effectively and interpret the data to generate meaningful implications (Dodge et al. 2015, Jayaprakash et al. 2014, Macfadyen & Dawson 2010, Smith et al. 2012, Wise 2014, Wise et al. 2014). Utilising such data for early identification of students who are at academic risk allows educators to provide timely and viable interventions that can help improve academic achievements and, ultimately, raise the quality of the university experience for the students.

Higher education institutions are increasingly adapting Virtual Learning Environments (VLEs) to support teaching and learning processes. VLEs are online learning platforms that deliver online educational objects in addition to providing the digital space to allow students to communicate with their peers and lecturers (Laister & Kober 2002). The term VLE is used as a general description of a range of electronic learning platforms which have been used as virtual educational tools in practice. These tools include, but are not limited to, learning and course management systems. Therefore, in this thesis, we refer to different types of electronic educational systems mentioned in the literature generically as VLE.

Generally, higher education institutions utilise learning management systems (LMSs) to manage their VLE. LMSs are software applications that facilitate e-learning components and allow teachers and students to interact with digital learning components (McGill & Klobas 2009). There are many well-known LMSs available such as *Moodle*, which is an open-source LMS and others which are commercially distributed, including *Blackboard* and *desire2learn*.

Integrating VLEs in the teaching and learning process provides many benefits, including delivering educational components, organising students' assessment activities and helping users to manage their learning activities (Coates et al. 2005). Utilising VLEs as an educational tool alongside traditional face-to-face teaching methods forms a blended learning mode (Garrison & Kanuka 2004). VLEs provide sets of tools to allow students and lecturers to establish synchronous and asynchronous communications, virtually without chronological or physical limitations (Coates et al. 2005, Loncar et al. 2014).

Despite the primary goal of the VLEs being to improve the learning experience, VLEs also record a very high volume of detailed data captured from students' interactions with various virtual learning objects, which can be analysed to optimise students' academic outcomes. Students' interactions with VLE components and contributions within discussion forums data may reflect their learning patterns, which can be used to identify at-risk students.

Many studies have been conducted to investigate the association between academic performance and various aspects of online learning activities which lead to deeper understandings of academic risk factors. For instance, past studies conducted by Agudo-Peregrina et al. (2014) and Cerezo et al. (2016) observe a positive correlation between multiple aspects of VLE and course performance. On an institutional level, the University of Maryland, Baltimore County, detected an association between students' achievements and VLE usage, whereby active users obtained higher grades (Fritz 2011). Furthermore, Sclater et al. (2016) highlights multiple studies which have found a significant positive relationship between students' VLE engagement levels and students' success. For example, the authors reviewed a study conducted at Nottingham Trent University where researchers found a strong relationship between students' academic achievements and levels of engagement.

Moreover, other efforts compared the predictive power of VLE predictors over personal factors. Sclater et al. (2016) report a study utilised data collected from California State University, where researchers indicate that predictors extracted from VLE activities data were four times more significant than demographic characteristics. Furthermore, many studies investigated digital traces data to identify the most inflectional factors associated with student risk behaviours and subsequently to predict students' retention rates (i.e., (Aguiar et al. 2014, Bayer et al. 2012, Chai & Gibson

2015)) and final achievements (i.e., (Conijn et al. 2017, Dascalu et al. 2016, Jishan et al. 2015, Villagra-Arnedo et al. 2017)) which enhance the university's ability to deliver interventions to at-risk students.

In 2017, there were over 1.5 million tertiary students studying Australia (*Department of Education and Training, Australia* 2018). Based on data obtained from the *Department of Education and Training, Australia* (2018), enrolments in Australian higher education institutions jumped by about 80 percent between 2001 and 2017. With the significant increase in the number of students attending university and the huge amount of student online-interactions data being collected, universities and lecturers have become increasingly interested in automated tools to analyse online learning behaviours and enhance learning experiences and outcomes accordingly.

In the educational research community, researchers refer to the phenomenon of analysing student-related data to evaluate learning progress, including those factors that influence university academic performance as Learning Analytics (LA). In other words, LA are concerned with the analysis of student-generated and digital-traces data, in order to monitor and report students' learning activities. Furthermore, LA analyse various aspects of student-related data, including academic and demographic factors, in order to improve academic outcomes.

A number of conflicting definitions have been proposed for LA. A definition suggested by the Society for Learning Analytics, conceptualises learning analytics as "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" (Siemens 2011). This rather broad definition underscores the complexity and multiplicity of learning analytics. In contrast, Educational Data Mining (EDM) is another research field that focuses on similar objectives in educational context (Aldowah et al. 2019). The EDM definition suggested by the International Educational Data Mining Society, defines the EDM as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in" (Siemens & Baker 2012). Throughout this thesis, we refer to both research fields as LA.

LA can assess a learner's degree of engagement within their social learning community using various analysis methods and techniques for student-related data. Digital learning data generated from students' interactions and engagements within VLEs are the main drivers of these analytics process approaches. LA methods allow institutes and educators to evaluate and optimise learning and teaching processes.

LA approaches have been used for a range of learning-related purposes, such as detecting students who are at-risk of failure or withdrawing from their course (Azcona & Casey 2015, Bainbridge et al. 2015, Bydzovska 2016, Hu et al. 2014, Jayaprakash et al. 2014, Smith et al. 2012). Furthermore, LA methods also have been utilised for identifying group performance (Cen et al. 2016), supporting selfregulated learning (Manso-Vzquez & Llamas-Nistal 2015), optimising collaborative learning experiences in social communities (Knutas et al. 2013) and understanding the structure of a small group of students to support them (Goggins et al. 2010). These applications enhance the higher education institute's ability to provide practical support by providing timely, direct and personalised interventions for students in need, or utilising corrective actions at course level by modifying the course structure to suit learning requirements.

Several studies employ LA approaches to monitor and predict students' learning performance, particularly by identifying students who might achieve poor final course achievements. Analytics approaches were used to analyse digital learning traces to extract students' characteristics. For example, Social Network Analysis (SNA) has been used as a branch of LA in many educational applications to understand, evaluate and improve the social structure within social learning communities. A number of studies utilised SNA to investigate the relationship between participation rates in online collaborative learning environment and learning performance (Cheng et al.

2011, Gunnarsson & Alterman 2012), while other researchers considered mapping the social structure via SNA to visualise students' collaboration patterns (Haig et al. 2013, Macfadyen & Dawson 2010). Moreover, many other studies have used students' social characteristics, extracted using SNA, as predictors of their performance (Haig et al. 2013, Macfadyen & Dawson 2010, Romero et al. 2013).

Moreover, numerous other aspects of students' online interactions patterns have been analysed to predict students' academic performance. Quantitative analysis approaches are the most popular methods used to investigate the students' degrees of engagement with VLE components (i.e. (Arnold & Pistilli 2012, Falkner & Falkner 2012, Haig et al. 2013, Rogers et al. 2014, Wolff et al. 2013)). On the other hand, only limited studies have considered performing qualitative analysis of students' engagement data with VLEs. For instance, in terms of analysing students' contributions in online discussion forums, typically researchers employ numerical analysis approaches to extract students' engagement factors including frequency analysis of creating posts, counting the number of words in the post (Dascalu et al. 2016, Lopez et al. 2012), and computing the duration of participation (Morris et al. 2005). A few studies have considered undertaking qualitative methods such as analysing posts' textual content to evaluate sentiments expressed in posts (Binali et al. 2009), observe students' opinions about the course structure and resources (Ashenafi et al. 2016) and detect learners' confusion in discussion forums (Yang et al. 2015).

Learning analytics utilises many predictive methods to predict future learning-related events such as course outcomes and identifying students who are willing to drop-out of courses. Several studies employ a range of statistical and analytical approaches to fulfil prediction tasks, including forecasting students' who are at academic risk based on a variety of predictors and data sources (Bainbridge et al. 2015, Bydzovska 2016, Howard et al. 2018, Hu et al. 2014, Jayaprakash et al. 2014, Shelton et al. 2017, Wolff et al. 2013). Many of the predictive approaches used in LA contexts are borrowed from other research fields, particularly machine learning.

Machine learning algorithms are commonly used in the field research of learning analytics. A wide variety of machine learning techniques have been used to predict future learning events including bayesian, decision tree, clustering, regression, neural network, support vector machine and rule-based approaches. The predictive power of machine learning approaches raises interest in utilising them to identify at-risk students in the early stages of the semester. Researchers and higher education institutions develop early warning systems of at-risk students with the help of machine learning approaches (Agnihotri & Ott 2014, Choi et al. 2018, Dodge et al. 2015, Dominguez et al. 2016, Howard et al. 2018, Hu et al. 2014, Jayaprakash et al. 2014, Macfadyen & Dawson 2010, Smith et al. 2012). Various data types, LA analysis approaches and prediction techniques have been used moving towards achieving early and quality detection of students academic risk. The purpose of these models is to allow early identification of and providing meaningful support for students in need, thereby optimising students' learning experiences and university outcomes.

Ensemble modelling is widely used in a machine learning setting to develop predictive instruments (Galar et al. 2012). However, the ensemble method is used rarely to predict learner's performance in higher education contexts (i.e. (Boyer & Veeramachaneni 2016, Er et al. 2017)). Ensemble-based models are constructed by combining multiple classification approaches to enhance the predictive accuracy over a single learning model (Dietterich 2000). Merging several classifiers in a single model allows us to utilise a collection of hypotheses from the hypothesis space, which can help to improve prediction quality by reducing the misclassification rate. In ensemble modelling, nominated classification members are ensembled by combining their outputs. Various mechanisms have been used to combine members' predictions to produce final prediction decisions, such as averaging members' probabilities to form the ensemble model's final output.

However, a key trend in past work is that researchers usually rely solely on quantitative analysis to extract learning characteristics from online learning activities data (i.e. (Cen et al. 2016, Chai & Gibson 2015, Conijn et al. 2017, Mueen et al. 2016, Pardo et al. 2016, Villagra-Arnedo et al. 2017)). Researchers rarely consider qualitative aspects of learning behaviours such as the change in learning behaviour patterns over time and the analysis of online contributions data generated by students such as post contents. Utilising such information may result in more accurate predictions, as they reflect an important part of learning characteristics.

Furthermore, predictive models are typically trained with and evaluated using data drawn from a single course (i.e. (Cassells 2018, Dominguez et al. 2016, Hayes et al. 2017, Howard et al. 2018, Hu et al. 2014, Jishan et al. 2015, Lopez et al. 2012, Romero et al. 2013)), Training a predictive model using data gained from a single course would make the model applicable for a single course where such training approach may lead to a predictive model performs poorly when tested with data drawn from other courses or subsequent semester (Ahadi et al. 2015). However, in other cases, researchers have employed data drawn from dierent courses to develop predictive models. This fact raises concerns about the scalability of these predictive models in dierent cultural or educational settings.

Another key observation is that predictive models existing in the literature are built in a static machine learning environment, where these models are fixed. In a static development environment, predictive models are unable to cope with any changes which may occur in the prediction space dynamically (Gama 2010) due to the absence of adaptive mechanisms. Adaptive methods allow underling predictive models to update their properties automatically, based on recent changes occurring in the prediction space (Gama 2010).

The particular focus of this thesis is to produce reliable, early predictions of students who are at academic risk, based on VLE interactions and participation data in blended learning setting. In such a learning environment, relying on online learning activities to predict students' performance alone is a challenging task, as students may perform off-line learning activities and use communication channels which reduce the need for online learning components. Consequently, they reduce the amount of learning activities data available for particular students and reduce the ability of the predictive instruments to distinguish actual classes.

Therefore, to achieve the highest possible reliable prediction results, we employ various quantitative, qualitative and social approaches to analyse students' VLE activities and contributions in online discussion forums. Moreover, the extracted learning variables are used to develop an exemplar multi-course early warning framework of students at academic risk in higher education settings. The proposed multi-courses early warning framework is developed using a novel Grey Zone ensemble model, proposed in this thesis to enhance the framework's ability to distinguish if an instance under prediction is actually at-risk. Furthermore, in order to enhance the scalability and adaptivity of the proposed multi-courses early warning framework and ensuring its ability to adapt and make continuous improvement dynamically, we developed and evaluated an adaptive framework applicable to the novel Grey Zone design proposed in this thesis.

In this study, online learning data was collected from thirteen blended computer science courses taught at the University of Adelaide, Australia, over the rst and second semesters between 2012 and 2016 (N = 1,476 enrolments). Gathered data was used to extract predictive features of student's academic risk by employing learning analytics methods including mining student-generated textual-based contents with the help of the CoreNLP toolkit. Sets of experiments were then conducted to predict students' academic risk on a weekly basis by utilising the predictive power of logistic regression approach. Finally, a variety of evaluation metrics were used to evaluate the performance of developed prediction instruments.

1.1 Research Questions

To support the primary objective of this thesis, which is concerned with detecting students who are at academic risk early in the semester based solely on online learning behaviours, this thesis addresses the following research questions:

- What are the most influential student online discussion forum participation predictors for students who are at-risk in a blended learning setting?
- What technology is needed to enhance the ability of the predictive model to produce reliable predictions of students who are at-risk?
- How can a reliable early warning framework of at-risk students that supports multiple courses be developed using VLE interactions and discussion forum data in a blended learning setting?

Furthermore, to support the second objective of this thesis, which is to allow the predictive instrument to cope with any changes which may arise in the prediction environment dynamically, this study addresses the following research question:

- What are the adaptive strategies that can be used to allow the proposed framework to cope with any changes that may occur in the prediction space dynamically to maintain its ability to produce reliable predictions?

1.2 Original Contribution

This thesis tackles issues regarding predicting students who are at academic risk early in the semester, using a variety of learning characteristics drawn from online learning behaviours in blended learning setting. The main significant contributions are listed as follows:

- Extracting a range of quantitative and qualitative characteristics from online learning activities data in blended learning setting. Extraction methods, including the development of an automated approach to score the strength of students' sentiments expressed in discussion forum posts alongside computing variables, reflect the changes in students' learning patterns over time. The sentiment weighting approach works by scoring the strength of students' sentiments as they are expressed in discussion forum posts, based on corresponding adverbs, by using a Digital Adverb Strength dictionary that has been developed for this purpose. The dictionary was stored in an XML format file to make it easier for other researchers to benefit from it. Moreover, to determine the influence of the extracted discussion forum predictors in the predictive model, the predictors were ranked based on their importance to the model using multiple well-known feature selection methods.

- Proposing a novel Grey Zone design to improve the performance of binary predictive models. The design aims to identify the range of probabilities where most of misclassifications occur (which can be considered the weakness point of a predictive model) and address this problem by utilising an alternative Grey Zone prediction model. The Grey Zone models are expert in distinguishing prediction classes of instances falling in the Grey Zone. The effectiveness of the proposed Grey Zone design was carefully evaluated by utilising the Grey Zone design to build a predictive model of at-risk students on a weekly basis, trained on discussion forum contributions only. The impact of the proposed Grey Zone design on the model's performance was evaluated against a traditional model design in terms of the overall prediction accuracy and Area Under the Curve (AUC) metric (Hanley & McNeil 1982). The results illustrated that Grey Zone modelling improved the model's performance significantly.

- Developing an exemplar multi-course early warning framework of at-risk students that utilises online learning activities and online discussion forum participation data to forecast students' course performance where the framework employs the novel Grey Zone design proposed in this thesis. The framework's development process involves constructing a fixed-size ensemble predictive model, where each model member is an expert in a local area of the features space. The performance of the multi-course early warning framework, uses an unseen dataset to examine the model's performance when it predicts future events. The evaluation of a fresh dataset belongs to four courses, where each course is different in terms of its online activities distribution. Moreover, by analysing the system's predictive power using fresh data, we identify the best week to provide interventions to students who have been identified as at-risk, where the framework starts to provide reliable, high-quality predictions.

- Developing and evaluating an Adaptive Grey Zone Ensemble Model (AGZEM) framework that is aligned with the Grey Zone design, which allows the multi-course framework to cope with any changes that may occur in the prediction environment. The development of an AGZEM framework involves building two complementary algorithms: an Ensemble Model Adaptive (EMA) algorithm and a Grey Zone Bounds Adjustment (GZBA) algorithm. The evaluation of the proposed adaptive framework is deployed with and without historical data forgetting criteria to examine adaptation outcomes over both scenarios.

1.3 A Guide to the Thesis

Chapter 2: Predicting Students' Academic Performance: A Review

Chapter 2 presents past efforts towards identifying significant predictors of students' academic outcomes related to academic, performance and demographic characteristics. Moreover, this chapter reviews in detail previous efforts to develop automated predictive models and early warning systems of students' success. Finally, it summarises various efforts in the literature to predict students' performance in blended and online learning settings.

Chapter 3: An Adaptive Multi-Course Early Warning Framework for At-Risk Students

Chapter 3 presents a high-level explanation of the work included in the thesis to bridge the gap identified in the literature. This chapter presents a general description of the VLE characteristics used in this study. Furthermore, it presents an explanation of the proposed novel Grey Zone design used in this thesis. Finally, it describes the developed early warning framework for at-risk students and the adaptive mechanisms used to update the framework dynamically.

Chapter 4: Context and Data

Chapter 4 provides a brief description of the data, methods and tools used in this thesis, alongside describing the utilised evaluation metric. This chapter presents data gathering and preparation processes and provides statistical descriptions of the collected dataset for thirteen blended learning courses, followed by a detailed description of all the variables utilised in different chapters of the thesis. Moreover, it presents a description of the developed Digital Adverb Strength dictionary and its development process.

Chapter 5: Early Detection of At-Risk Students Using Course Discussion Forum Data

The first half of chapter 5 presents a detailed description of the proposed automated process of the weighted sentiment approach used to evaluate the strength of students' posts posted in course online forums and employs the Digital Adverb Strength dictionary described in Chapter 4. Moreover, it presents in detail the proposed novel Grey Zone strategy for binary classification to enhance the binary classifiers' ability to produce accurate predictions. The rest of the chapter evaluates the degree of importance of each extracted discussion forum feature to determine the most influential predictors. Furthermore, it experiments with the predictive power of the extracted features by using them to develop predictive models, alongside experimenting with the impact of Grey Zone modelling on the weekly models' performance.

Chapter 6: Exemplar Multi-Course Early Warning Framework to Identify At-Risk Students in Blended Learning Computer Science Courses

Chapter 6 extends the features coverage by utilising additional online learning activities to develop an exemplar multi-course early warning framework of students who are potentially at academic risk. This chapter begins by describing the fixed-size ensemble modelling utilised to construct underlying predictive models, where the framework follows the proposed novel Grey Zone strategy. It presents an experimental study conducted to evaluate the quality and accuracy of the framework's predictions for future events using an unseen evaluative dataset.

Chapter 7: Towards an Adaptive Early Warning Framework for At-Risk Students

Chapter 7 sheds some light on popular adaptive mechanisms in the literature used to update predictive ensemble modellings. Moreover, it presents additional adaptive strategies applicable to the Grey Zone design proposed earlier. Adaptive mechanisms are powerful tools to cope with changes in the predictions space as a result of changes in the educational or cultural settings, or changes in students' learning behaviours over time. The chapter describes details of the Adaptive Grey Zone Ensemble Model (AGZEM) framework, the associated Ensemble Model Adaptive (EMA) algorithm and the Grey Zone Bounds Adjustment (GZBA) algorithm which are the vehicle for the experiments in this chapter. The experimental study performed in this chapter examined the impact of the proposed framework and algorithms on the predictive models' quality over two scenarios. The scenarios involve appending adaption batches to existing ones and utilising a forgetting mechanism for the historical data.

Chapter 8: Conclusion and Future Directions

Chapter 8 presents an overview of the contributions and limitations of the thesis alongside describes possible future work and closing remarks.

Chapter 2

Predicting Students' Academic Performance: A Review

This chapter presents past efforts towards identifying students' academic outcomes and the effects of modern technology on evaluating learning performance in the higher education context. Furthermore, it sheds lights on different academic, learning and demographic characteristics that have been identified in the literature as predictors of students' academic outcomes. Moreover, this chapter reviews previous efforts towards developing predictive models and early warning systems of students' success, along with evaluating the impact of these instruments on students' academic achievements at an institutional level.

This chapter is organised as follows: Section 2.1 presents a profile regarding the developments in identifying tertiary students' performance. Then, Section 2.2 discusses the different types of student performance predictors utilised in the literature in terms of data source and extraction approaches. Section 2.3 reviews various efforts and case studies conducted to develop predictive models in higher education settings. The latter section presents several efforts towards developing early warning systems for students who are at academic risk and their effectiveness in improving academic outcomes. Finally, Section 2.4 summarises various studies in the higher education literature to identify the gaps and outlines how this thesis addresses these gaps.

2.1 Overview

Interest in evaluating learning processes in higher education institutes has risen for over six decades. Pantages & Creedon (1978) reviewed studies conducted to investigate factors associated with attrition and academic success in higher education between 1950 and 1975. Since that time, considerable effort has been conducted to identify college student academic retention and outcomes characteristics. This research enables universities to improve their educational and outcomes quality, as well as reducing drop-out rates.

In the 1970s, Spady (1970), Terenzini & Pascarella (1978) and Tinto (1975) investigated factors leading students to fail to complete their academic courses in the higher education context. In 1984, Astin (1984) also studied the relationship between university students' physical and academic efforts, and their decision to drop-out of courses, where he identified characteristics and used them to develop a student retention model based on students' physical and pedagogical involvement.

Decades later, a new style of education emerged. In the 1990s, many higher education institutes transformed from depending primarily on traditional modes of learning and teaching to adapting advances in technology for teaching and learning processes, by means such as employing modern media technologies in distance education programs. Additionally, in order to provide supplementary educational options, some universities have shifted further to deliver online programs. For example, in 1993, Jones International University offered the first accredited fully online courses, followed by the California Virtual Campus in 1997 and the British Open University in 1999 (Casey 2008).

In the twenty-first century, with the digital and ICT revolution, numerous universities have gradually merged virtual learning environments with on-campus courses or even come to offer fully online courses (Bates 2005). VLEs use the internet to deliver educational components, which then provide new communication and collaboration media and allow students to submit assignments online and undertake online exams. The digital trace generated from students' online interactions with virtual learning platforms offers opportunities to explore and reach deeper understanding of student learning patterns and needs (Wolfgang & Hendrik 2012). Moreover, digital footprints can be captured to enable universities to identify the most inflectional factors of students' learning performance, subsequently identifying poorly performing students (i.e., (Cassells 2018, Falkner & Falkner 2012, Jayaprakash et al. 2014, Wolff et al. 2013)), predicting students' retention rates (i.e., (Aguiar et al. 2014, Bayer et al. 2012, Chai & Gibson 2015)) and final achievements (i.e., (Conijn et al. 2017, Dascalu et al. 2016, Jishan et al. 2015, Natek & Zwilling 2014, Villagra-Arnedo et al. 2017)), alongside visualising online learning patterns (i.e., (Haig et al. 2013)).

While earlier LA were perceived as limited, primarily to the issue of identification of at-risk students (i.e., (Rogers et al. 2014)), the dramatic growth in the number of students in tertiary education and the increasing complexity of learning courses gave rise to a number of new modes of implementing learning analytics in educational environments. LA can be beneficial for different stakeholder levels. For instance, at a student level, LA may be used to encourage self-reflection, and to provide performance assessment and feedback (Almosallam & Ouertani 2014). Moreover, it may help lecturers and institutions to reform course structures based on learning requirements, monitor learning behaviours, as well as predict students' academic performance and provide interventions to those in need (Almosallam & Ouertani 2014).

In fact, data plays a key role in the investigation of learning processes. It is important to select data that suits each analytical and prediction technique, keeping in mind the amount and nature of the data. Proper data selection can significantly affect the suitability and validity of the results (Hernández-García & Conde 2014). Various types of personal and educational data have been used to analyse students' learning patterns and assess students' performance in higher education settings, including data generated from actions performed on social learning platforms, demographic and pre-enrolment factors, and academic performance records.

In the last decade, many studies employed the power of LA on their students' data to improve academic outcomes. Higher education institutions applied data extracted from multiple sources using LA approaches to achieve maximum benefit. Muthukrishnan et al. (2017) highlight past efforts on developing data-driven predictive models to predict students' retention and academic performance using a variety of student variables. Furthermore, Bin Mat et al. (2013) and Sclater et al. (2016) reviewed a range of LA-powered tools that have been developed and used at various higher education institutes in Australia, Germany, the UK and the US. The reviewed tools focus mainly on improving students' academic achievements where they offer various sets of functionalities including early identification of at-risk students, providing personalised interventions to students and visualising students' social learning behaviours.

On the other hand, Leitner et al. (2017) discuss concerns about the scalability of existing LA predictive models, where they may be applicable only within the same educational and cultural setting. Scalability concerns have arisen since existing models are developed using data drawn from a single educational institute and have never been tested using instances obtained from different educational or cultural contexts such as in (Aguiar et al. 2014, Ashenafi et al. 2015, 2016, Azcona & Casey 2015, Bayer et al. 2012, Bydzovska 2016, Lopez et al. 2012, Natek & Zwilling 2014, Shelton et al. 2017).

LA predictive instruments are built upon a set of predictors to indicate numerous aspects of learning progress, including academic risk status. Predictors are usually extracted from diverse student-related data drawn from one or multiple sources. The following section describes different sources of predictive variables used in the literature for predicting students' academic performance.

2.2 Academic Performance Predictive Variables

In higher education settings, many researchers exploit data drawn from various digital data warehouses to build predictive models, including records of students' static, dynamic and academic factors. However, others employ an alternative data collection methodology by conducting surveys to gain knowledge of students' personal characteristics (Gray et al. 2014) or by questioning students about their learning experiences (Pardo et al. 2016, Sorour et al. 2016). While several models are limited to only one type of data (i.e., (Dascalu et al. 2016, Hu et al. 2014, Macfadyen & Dawson 2010, Romero et al. 2013)), the majority mix data from various sources (i.e., (Aguiar et al. 2014, Bayer et al. 2012, Chai & Gibson 2015, Conijn et al. 2017, Howard et al. 2018, Jayaprakash et al. 2014, Mueen et al. 2016, Shelton et al. 2016, Wolff et al. 2013)) to build predictive models. Students' data can be grouped into three main categories, according to sources which collect static, dynamic and academic data.

Static data involves students' demographic characteristics and personal elements. The demographic factors include personal information such as gender, age, race, economic background and pre-university performance. Static predictors were most commonly used in early versions of student attrition models; although they have been used also in a number of more recent efforts to predict university outcomes and support students to achieve their academic goals (i.e., (Aguiar et al. 2014, Arnold & Pistilli 2012, Rogers et al. 2014, Wolff et al. 2013)). Rogers et al. (2014) relied heavily on static and academic factors in their framework, which detects students who are at-risk of academic failure, while, Wolff et al. (2013) integrated demographic data with data drawn from other sources to identify at-risk students. However, despite the value of the dynamic data, Wolff et al. (2013) observed that static data had no significant impact on the outcome of the prediction instrument. Furthermore, in another study, researchers reported that dynamic factors are four times more powerful in predicting students' success than static characteristics (Sclater et al. 2016).

In social learning settings, dynamic data is defined as the data recorded about learners' interactions with VLE elements and online activities throughout the learning processes (Rogers et al. 2014). These data are usually stored automatically in special data warehouses. The widespread nature of VLE adaptation allows for the recording of significant volumes of online educational data. There are two main types of dynamic data, according to the nature of the data. The first type is VLE interactions data, where each data entry is a record of an action performed in online learning platforms such as logins, views, and interactions with online learning components. VLE interactions dynamic data are widely used in learning analytics tools (i.e. (Arnold & Pistilli 2012, Falkner & Falkner 2012, Haig et al. 2013, Rogers et al. 2014, Wolff et al. 2013)). The second type is online contributions data, which contains user-generated textual contents and instances of social presence, such as posts, text chats and social engagement within a social learning environment. Various online contributions data have been used in a number of learning analytics-based tools (i.e. (Caballe et al. 2011, Ferguson & Buckingham Shum 2011, Ferguson et al. 2013, Rabbany k. et al. 2012)).

The third group of university students' data contains academic elements. Academic data includes records of students' learning behaviours, post-enrolment performance and their previous academic history. The application of such data may lead to a deeper understanding of students' risk patterns as they reflect students' preliminary course performance as well as overall academic degree achievements. A few projects utilise this type of data to support students. Mueen et al. (2016) predict which students are in danger of failing their courses by utilising a mix of students' current performances and past academic records with other static data.Arnold & Pistilli (2012) pair students' academic history data with other types of data to identify at-risk students. Jishan et al. (2015) predict students' grades based on CGPA, interim assessments and attendance records.

Previous studies paid significant attention to identifying the most powerful predictors of student outcomes. For example, Natek & Zwilling (2014) explored the correlation between various student variables and their final grades in higher education settings to determine key influential characteristics. Many of these studies point out the powerful influence of VLE predictors in forecasting students' academic performance (i.e., (Agudo-Peregrina et al. 2014, Cerezo et al. 2016, Sclater et al. 2016, Wang et al. 2001)). Furthermore, due to the valuable predictive influence of variables extracted from online learning activities in predicting students' success, a number of studies count solely on dynamic features to develop predictive models such as in (Dascalu et al. 2016, Hayes et al. 2017, Hu et al. 2014, Lopez et al. 2012, Romero et al. 2013, Xing et al. 2015). The rest of this section discusses various forms of predictors drawn from VLE activities and contributions used to predict students' academic performance.

2.2.1 VLE Interactions Predictors

Learners' VLE interactions data enable the tracking of students' online learning behaviours and the recognition of students' learning patterns. Furthermore, they reflect learners' degrees of engagement with different online learning objects. VLE logs are a rich source of students' interactions with learning components within online learning platforms. The logged data includes, but is not limited to, data about the frequency of login-in to VLE, accessing learning objects, submitting and re-submitting assignments and the number of online exam and quiz attempts. Furthermore, each logged event is associated with other detailed information, such as the event's date and time, user details and IP address.

Several studies examined the correlation between a number of VLE logged activities and students' performance in online and blended learning setting to identify the most significant predictors. Agudo-Peregrina et al. (2014) studied eight blended and fully online courses and they observed a relationship between student online behaviours alongside other factors and their final performance in online learning models, however, the correlation was not significant in the other learning model. Another study conducted by Kim et al. (2014) shows that student final performance can be predicted through their online activities in the blended learning setting. Furthermore, Cerezo et al. (2016) examined the correlation between students' online participation in VLE-supported courses and their final achievements. The authors conclude that some variables are correlated positively with learners' final achievements.

Although VLE interactions are the most widely used source of dynamic predictors (Aguiar et al. 2014, Azcona & Casey 2015, Cen et al. 2016, Chai & Gibson 2015, Conijn et al. 2017, Pardo et al. 2016, Villagra-Arnedo et al. 2017), they present only one side of the dynamic data. Online discussion forum data can reflect valuable aspects about students' learning progress, which can lead to better understanding of students' learning factors.

2.2.2 Online Discussion Forum Predictors

Online discussion forums and other online communication tools have become essential elements of social learning settings, which provide a virtual space for students to seek help, express their concerns, share information and learn from others. Web-based discussion boards allow students to interact with educators and peers within the social learning environment. Various studies identify the relationship between online forum participation and learning outcomes. For example, Cheng et al. (2011) detect a positive correlation between student involvement in online discussion boards and learning performance. In another work conducted by Shaw (2012), he observed a trend between the active use of online forums and learning performance in an online programming language course. Furthermore, other study indicated that the participation of students in online communication tools helps them to succeed academically (Gunnarsson & Alterman 2012).

In the higher education context, exciting prediction-based models that employ data driven from online discussion forums typically rely on quantitative analysis of students' contributions. A common practice is to analyse student involvement by counting the frequency of actions performed and amount of time spent on the online forums (Xu et al. 2016). Several studies have performed numerical measurements of discussion media interactions and post contents data, such as the frequency of posting messages and counting the number of words in the post to predict students' performance (Caballe et al. 2011, Dascalu et al. 2016, Lopez et al. 2012), while others have combined the same types of variables with additional factors (Bainbridge et al. 2015).

On the other hand, other researchers have considered additional aspects of online forum data, such as mapping social structures via social network analysis, to visualise students' collaboration patterns and forecast learners' final achievements (i.e., (Haig et al. 2013, Lopez et al. 2012, Macfadyen & Dawson 2010)). Additionally, others have explored posts' textual contents manually such as (Caballe et al. 2011, Romero et al. 2013) or by utilising automated natural language processing approaches (Adamopoulos 2013, Wen et al. 2014*a*, Yang et al. 2015) to evaluate the qualities and purposes of each communication. For instance, in a blended learning context, Romero et al. (2013) propose a predictive model to forecast student risk behaviours based on qualitative, social network and VLE interactions information extracted from students' involvements in online discussion forums. The authors processed the qualitative information by inviting the lecturers to score the content of the messages manually, alongside investigating social network aspects obtained from social learning analysis. Moreover, Adamopoulos (2013) applied textual analysis to online contributions data to develop explanatory and predictive models of students' completion in MOOCs. Furthermore, (Yang et al. 2015) conducted research to detect learners' confusion based on analysing the contents of forum posts in Algebra and Microeconomics MOOCs, then they examined their influence on student retention. The study revealed that there is a statistical correlation between drop-out rates and confusion factors.

Social Network Analysis

While social network analysis has its roots in sociology, it has been used recently as a branch of learning analytics (Buckingham Shum & Ferguson 2012, Filvà et al. 2014, Rabbany k. et al. 2012). A social network analysis studies a set of social actors or network members, as well as their interactions, relationships and contributions (Knutas et al. 2013, Liu 2011, Rabbany k. et al. 2012). A social network analysis represents network features numerically or visually in order to analyse them quantitatively or qualitatively (Rahman & Dron 2012). To describe a social network visually, interactions and relationships have to be mapped into a communication matrix that can be visualised (Knutas et al. 2013). An alternative approach is to use graph theory to understand a social network in a quantitative manner (Filvà et al. 2014). In graph theory, each social actor is represented by a vertex (node) and each communication link is presented as a relationship. The popularity of each node is measured by the node degree, which is the number of edges from the community connected to that node (Rabbany k. et al. 2012).

The growing awareness of the importance of social network analysis in educational environments has led many researchers to apply it in their own research. Numerous researchers have used a social network analysis to identify students' collaboration and communication patterns within social learning settings. Haig et al. (2013) and Macfadyen & Dawson (2010), all employed social network analysis alongside other approaches to monitor and identify student behaviour patterns within learning management systems. They also developed frameworks to predict which students were at-risk at an early stage. Knutas et al. (2013) analysed communication patterns in a collaborative course using a social network analysis method in order to understand and optimise the collaborative learning process. Goggins et al. (2010) used social learning analysis to understand the structure of a small group of students in order to support them in an online course.

Several metrics derivative from SNA have been used to predict student success. Romero et al. (2013) pointed out that the degree of centrality and the degree of prestige are the most significant SNA predictors of university student performance. In a MOOC setting, Jiang et al. (2014) detected a positive link between the social network degree captured from the first week data and students' later performance.

Content-based Analysis

Analysing user-generated content in discussion forums opens doors to exploring some hidden dimensions of the learning experience, reflecting learners' motivations and experiences. In an educational context, numerous studies have taken advantage of automated techniques that organise and classify textual content and analyse them. Among those approaches, NLP, especially sentiment analysis, are the most commonly used techniques for the prediction of student performance. NLP aims to evaluate and understand human-generated texts' linguistic properties automatically. NLP have been used to analyse student-generated textual content such as forum posts, however, the majority of the work has been applied within fully online learning and MOOC environments. NLP has been applied in a number of studies to monitor and enhance learning by detecting students' emotions in e-learning platforms (Binali et al. 2009, El-Halees 2011).

Some studies have explored the correlation between natural-language based variables extracted from participation in social learning environments and students' achievements. For example, Tucker et al. (2014) examined the relationship between the content of students' posts and their final outcomes in an art MOOC. Authors reported a minor positive correlation between posts and comments related to assessment and a strong negative connection with posts about specific assignments. Wen et al. (2014*a*) utilised sentiment analysis on online posts in a MOOC to observe students' feelings about the course. Researchers found a link between sentiment variables used in the study and learners' drop-out rates. Wen et al. (2014*b*) extended their work to detect students' opinions toward course structure and materials as expressed in forum posts, where they observed a significant correlation between extracted linguistic features and course completion rates.

Moreover, Crossley et al. (2016) and Robinson et al. (2016) mixed linguistic-based features with other demographic or online interaction features to evaluate their impact on predictive models over background-only or activity-only models. Crossley et al. (2016) evaluated the prediction power of features extracted from the language of posts and clickstream data to forecast the final achievements of 320 enrolments in a MOOC where linguistic-based features were extracted with the help of multiple automated NLP tools. However, the authors observed that although linguistic variables were predictive, activity-based features were the most powerful. On the other hand, aggregating both types of data improved overall prediction accuracy by about 10 percent. Robinson et al. (2016) employed NLP methods on pre-course open-response surveys that covered the students' intentions and course materials information from 27 MOOCs to predict students' intentions to complete the course. Combining language-based and demographic-based features enhanced the model's ability to predict outcomes when compared with the performance of another model which relied solely on static variables.

Finally, considerable efforts have gone towards identifying qualitative factors that can indicate academic risk in higher education. Various learning and personal sources of data were examined to recognise the most powerful characteristics. Dynamic data presents a quality source of predictors, where it illustrated it's predictive power against other data sources on several occasions (Sclater et al. 2016, Wolff et al. 2013). On the other hand, the main research attention has gone into performing numerical analysis of interactions data, with much less effort going into considering the analysis of online contributions dynamic data to extract features of students' success.

Various types of students' features have been used in many studies to construct predictive models of academic achievement with the help of a wide range of prediction methods. In the higher education literature, predictive instruments are varied in terms of the sources of utilised features, prediction outcomes formats and prediction approaches, but they all move towards the same target, which is to identify poorly performing students.

2.3 Student Success Predictive Instruments

The field of learning analytics is considered to be a new domain for predicting learners' performance and identifying at-risk behaviours. In recent years, there has been a trend to use automatic analyses and predictive approaches to better understand learning patterns and, ultimately, optimise learning outcomes. Along the research line, studies have commonly focused on developing and evaluating predictive models based on a dataset collected solely from a single course or multiple courses offered at a single institution. Numerous well-known prediction approaches were evaluated using various unities of information to determine the most accurate set of predictors such as in (Aguiar et al. 2014, Azcona & Casey 2015, Lopez et al. 2012, Mueen et al. 2016, Muthukrishnan et al. 2017).

Various student achievement predictive models were developed targeting diverse forms of prediction outputs including predicting at-risk/successful students, drop-out students, assignment scores or final grades. Several student-related data types have been fed into numerous approaches to build predictive models. However, classifications and regression techniques are the most popular approaches to forecast student academic achievement in the context of higher education. Among the utilised predictive approaches, Several studies stated that, by comparing multiple popular approaches, regression techniques produce the most accurate predictive results (Aguiar et al. 2014, Chai & Gibson 2015, Jayaprakash et al. 2014).

In statistics, a regression analysis is defined as a process involving several techniques for forecasting the relationship between the response (dependent) variable and a single or multiple explanatory (independent) variable(s). A regression analysis commonly provides valuable estimations. Consequently, it is one of the most regularly used prediction techniques in many scientific fields. Nevertheless, this prediction technique produces more reliable estimations when it deals with small numbers of variables, and big amounts of data, where changes are larger and more predictable and there are strong causal relationships. However, in some special circumstances, regression analyses may result in incorrect estimations.

Regression approaches are powerful techniques in predicting binary outcomes. For instance, Jayaprakash et al. (2014) selected a logistic regression technique to develop an early alert system to predict academic risk since an approaches comparison resulted in its identification as the best performing model in predicting students' risk status. In another study, Chai & Gibson (2015) compared different classification and regression algorithms to determine the most accurate approach to predict student retention. In their work, logistic regression obtained the most accurate results. On the other hand, several regression algorithms have also been used to predict more fine-grained targets. For example, Ashenafi et al. (2016) utilised a linear regression model to forecast final exam grades.

Early identification of students who are in danger helps instructors to provide timely interventions to students in need. While many researchers have used a semester aggregated dataset to evaluate their proposed prediction model (i.e. (Aguiar et al. 2014, Dascalu et al. 2016, Jishan et al. 2015, Mueen et al. 2016, Natek & Zwilling 2014)), others have examined a prediction model's performance in a timely manner to provide early results (i.e. (Ashenafi et al. 2016, Azcona & Casey 2015, Chai & Gibson 2015, Hu et al. 2014, Macfadyen & Dawson 2010, Pardo et al. 2016, Shelton et al. 2017)). Howard et al. (2018) proposed an early warning system to forecast students' final achievements on a weekly basis. Moreover, other models created by Pardo et al. (2016), Conijn et al. (2017) and Shelton et al. (2016) provide week-by-week results to predict students' final performance.

On an institutional level, Bin Mat et al. (2013) and Sclater et al. (2016) reviewed numerous early warning systems of academic performance used in various international higher education organizations, where most of the early warning systems are accompanied by some kind of intervention strategy for stakeholders. Moreover, several studies connect their predictive models with actionable strategies. For example,Jayaprakash et al. (2014) present the Open Academic Analytics Initiative (OAAI) which serves as an early alert system to deliver proactive interventions for students at academic risk. The OAAI program results in overall improvements in participating students' grades.

The rest of this section presents a review of the prediction models in the area of predicting students' performance in a higher education setting. Then it presents various attempts to employ early warning systems in higher education institutions and reviews the effectiveness of utilising them in terms of improving students' academic outcomes.

2.3.1 Prediction Models

When it comes to computational models to forecast students' final outcomes, the current literature mainly pays attention to comparing the performance of popular predictive approaches to identify the most accurate and reliable techniques. To compare prediction algorithms' predictive power, researchers generally employ identical set predictors on all underlying algorithms instead of selecting the most appropriate subset of predictors that suits each algorithm.

A generic method to forecast academic performance is by predicting courses' final outcomes, such as predicting students' likelihood to complete courses successfully. Many models have focused on predicting students' academic risk status (i.e. (Azcona & Casey 2015, Bainbridge et al. 2015, Howard et al. 2018, Hu et al. 2014, Jayaprakash et al. 2014, Shelton et al. 2017, Wolff et al. 2013, Xing et al. 2015)), while other models target more fine-grained predictive results by predicting students' individual assignments, exams or final course scores or grade (i.e. (Conijn et al. 2017, Dascalu et al. 2016, Jishan et al. 2015, Lopez et al. 2012, Natek & Zwilling 2014)).

Machine learning is the most commonly used approach to build predictive models in higher education settings. A wide variety of machine learning techniques, including Bayesian techniques, decision trees, clustering, regression, neural networks, support vector machines, and rule-based approaches, are compared and used to develop predictive models. Generally, among machine learning approaches, regression algorithms are the most popular approaches in the field of predicting students' academic performance (i.e. (Ashenafi et al. 2015, 2016, Bainbridge et al. 2015, Chai & Gibson 2015, Conijn et al. 2017, Dascalu et al. 2016, Jayaprakash et al. 2014, Rogers et al. 2014)), followed by decision tree algorithms (i.e. (Azcona & Casey 2015, Cen et al. 2016, Natek & Zwilling 2014, Pardo et al. 2016, Shelton et al. 2016, 2017)).

Multiple studies have compared the performance of regression algorithms against many other approaches. These comparisons aim to identify the most powerful prediction approach to build predictive models of student academic outcomes. For example, in studies conducted by Conijn et al. (2017) and Jayaprakash et al. (2014), students' variables extracted from various data sources were fed into multiple classification and regression techniques to predict students who were at academic risk. While Conijn et al. (2017) ran different regression-based algorithms using data captured from VLE logs and internal assessment scores,Jayaprakash et al. (2014), applied demographic factors and past performance data alongside online behaviours extracted from a Collaboration and Learning Environment to logistic regression, support vector machine, naive bayes and decision trees to compare their prediction accuracy. The results showed that a logistic regression algorithm generates the highest quality predictions.

Additionally, Aguiar et al. (2014) and Chai & Gibson (2015) compared various predictive techniques including random forest, logistic regression and decision trees algorithms using data collected from freshman courses to predict students who were at-risk of attrition. Aguiar et al. (2014) applied students' demographic, pre-enrolment, post-enrolment and electronic portfolio engagement data to compare the approaches, while Chai & Gibson (2015) used a wide range of static and academic variables to cover students' demographic, social, psychological, financial, enrolment and academic factors alongside VLE interactions variables. Both studies concluded that logistic regression models provide the most accurate results.
Although various studies elected regression approaches to develop binary prediction

models, other regression techniques can produce more concentrated prediction targets such as final grades and exams marks. In research conducted by Ashenafi et al. (2015), a linear regression model was trained on data to reflect a range of students' activities to forecast final exam grades in two courses. Students' data was collected from a web-based peer-assessment system implemented by researchers for an eight-week long course. A predictive model was designed to estimate the students' final exam scores on a weekly basis throughout the course period where a Root Mean Squared Error (RMSE) metric was used to evaluate the prediction quality. The model obtained an RMSE, in the final week of study, of 2.93 and 3.44 respectively for the two courses in predicting the students' final exam scores. Then, in a subsequent study in 2016, the authors intended to enhance the accuracy of the model using the same collected population (Ashenafi et al. 2016). The authors employed an alternative predictions strategy, a linear regression model that was trained for each study week using a subset of data that covered the period from the beginning of the semester up to the prediction week to train each model instead of using the full dataset. The main purpose of the proposed model training mechanism was to improve the prediction accuracy for successive weeks. In the second course, the prediction errors gradually reduced for successive weeks. In the first course, however, the RMSE decreased in the early weeks, followed by rises and then a slight decrease in the last week.

Dascalu et al. (2016) used a social media environment as a collaboration and communication tool in the context of a project-based learning (PBL) scenario, where a correlation was detected between the posts' contents and the students' final grades. The analysis revealed that there is a relationship between academic performance and the value of word entropy and number of verbs, prepositions, adverbs, and pronouns used. The authors conducted a stepwise discriminant function analysis using three features resulting from a linguistic-based analysis to predict whether students were under-performing or were in good academic standing.

On the other hand, other efforts in the area of developing computational predictive models of academic performance show that decision tree algorithms are also capable of producing more accurate classifications, when compared with numerous other prediction techniques. For instance, Azcona & Casey (2015) and Hu et al. (2014) compared a variety of approaches in terms of their predictive power using students' data collected from single or multiple sources. Both studies rely on variables extracted from VLE logs to train the models. In these studies, varied decision tree algorithms were compared against different sets of classifiers. Azcona & Casey's (2015) evaluation set contains a decision tree, linear regression, logistic regression, naive bayes, support vector machine and k-neighbours classifiers. Hu et al. (2014) compared a set of classification techniques composed of a decision tree, logistic regression and adaptive boosting. Sets of experiments were carried out to evaluate the performance of each individual approach and assemble the approaches into one predictive model. Azcona & Casey's (2015) results indicate that a decision tree is the best performing model, while Hu et al.'s (2014) evaluation of their results showed that CART tree-based approach accompanied by an adaptive boosting technique produced the most accurate predictive results.

Furthermore, Pardo et al. (2016) and Natek & Zwilling (2014) employed decision tree algorithms to build their predictive models. Pardo et al. (2016) used models to forecast midterm and final exam scores. Each predictive model corresponded to a lesson week throughout a thirteen-week course. The models trained on first year engineering students' online interactions data and internal assessment scores were derived from online learning resources, followed by assessment tasks where the models produced predictions on a week-by-week basis. In the other study, Natek & Zwilling (2014) targeted predicting students' final grades in higher education settings. They utilised selected, student-specific demographic, enrolment, and assessment factors to train a decision tree algorithm to predict the students' final grades.

In another example, Shelton et al. (2016) conducted a study to predict students' final achievements on a weekly basis throughout a sixteen-week semester. They collected data regarding students enrolled in twelve asynchronous fully online courses.

The dataset consisted of demographic and online activities variables. A time-series clustering analysis was applied to the dynamic variables, which then resulted in information which was combined with static variables. The resultant data was utilised to assess the predictive performance of six approaches: a decision tree, gradient boosting, rule induction, stepwise regression, forward regression, and backward regression. The decision tree models produced the most accurate classification results. The model identified up to 78.6 percent of at-risk students correctly, however, the model only started to provide reliable results at the tenth week and then continued to improve slightly until the end of the semester. In 2017, the authors continued working on an expanded dataset, aiming to produce more accurate and earlier predictive results by considering the variances in different learning patterns and course learning effort requirements (Shelton et al. 2017). The six proposed models were re-compared after employing the additional proposed features. The decision tree model remained the best classification model; however, the extra values resulted in earlier and more accurate predictions. By week six, the model was able to provide reliable predictions. The classifiers obtained a decent quality by achieving an overall accuracy of 89.26 percent with 85.45 percent of at-risk students being classified successfully.

Furthermore, other studies have stated that other machine learning approaches, such as naive bayes, support vector machines and clustering can provide accurate classifications of students' academic performance. Jishan et al. (2015) developed a decision tree, naive bayes and neural network models, which they designed to predict students' grades based on their CGPA, in-between assessments and attendance records. The authors applied Optimal Equal Width Binning and Synthetic Minority Over-Sampling data pre-processing techniques to treat the error rate resulting from imbalanced distribution of target classes in the training dataset. The model's evaluation results show that neural networks and naive bayes models outperform decision tree models where both models provide similar levels of accuracy.

Moreover, in work undertaken by Osmanbegovic & Suljic (2012) and Mueen et al. (2016) to forecast students' final achievements, the predictive power of naive bayes, neural networks and decision tree models were measured and compared. Both studies trained their models with students' demographic factors and past academic performance, but Mueen et al. (2016) included interim assessment scores in the training dataset. The naive bayes technique outperformed the other two techniques in both studies.

Nevertheless, others employ substitute approaches rather than traditional machine learning classification algorithms. In research carried out by Lopez et al. (2012), the authors examined the performance of various clustering and classification approaches to predict students' final achievements. Firstly, they analysed the students' participation in online discussion forum data to investigate the relationship between this and their final achievements. Features were extracted using a purely quantitative analysis of students' participation data in the forum. The results analysis revealed that by using a subset of selected attributes, including the number of posted messages and counts of words, the models produced the highest accuracy. With regard to model performance comparisons, the naive bayes classifier recorded the highest classification accuracy among the examined classifiers. The Expectation-Maximisation (EM) clustering algorithm achieved the best performance among the proposed clustering techniques.

Furthermore, Bydzovska (2016) worked on predicting students' final grades at the beginning of the semester. This study presented two approaches. The first approach searched for patterns in students' demographic and social behaviour data using classification and regression algorithms. To examine the first approach, students' historical data were used to train several classifiers including support vector machine, random forest, rule-based classifier, decision tree and naive bayes. The support vector machine classifier shows the best performance. Classifier accuracy was improved by combining students' social behaviour data with historical variables. The second approach was based on collaborative filtering techniques, where similar students' previous grades were used to predict the final grade.

Unlike the majority of the research in the field, Xing et al. (2015) proposed using genetic programming (GP) to predict students' performance. They developed a GP predictive model technique using Interpretable Classification Rule Mining from students' online activity data. Xing and his colleagues collected data about 122 students enrolled in one course. Also, the researchers considered building a number of machine learning models using same population. A performance comparison of GP against various machine learning and regression models was conducted to determine the most powerful prediction approach. The GP model was the best performing model in predicting students' performance by classifying 80.2 percent of students correctly.

2.3.2 Early Warning Systems of students' success

The importance of early identification of academic risk in higher education has led many researchers and institutions to invest in developing early predictive models of students' academic achievements. However, there is conflict in the descriptions of how early and how frequently early predictive instruments should produce outcomes. In the majority of cases, the models started reporting predictions on weekly basis (Ashenafi et al. 2015, 2016, Hayes et al. 2017, Howard et al. 2018, Macfadyen & Dawson 2010, Shelton et al. 2016, 2017). Others utilised an alternative procedure whereby the designed system reported at-risk students twice a semester, with the first results produced in the middle of the semester (Romero et al. 2013), or identifying the students' performance when every quarter of the semester had been completed (Jayaprakash et al. 2014). Furthermore, other efforts employed a pre-defined milestone to schedule prediction timings (Hu et al. 2014) or report students' risk status, based on assessment due dates (Hlosta et al. 2017).

Numerous studies have been conducted aiming to build timely and quality prediction models (Arnold & Pistilli 2012, Ashenafi et al. 2015, 2016, Cassells 2018, Conijn et al. 2017, Macfadyen & Dawson 2010, Villagra-Arnedo et al. 2017) which can serve as early warning systems of students' performance (Howard et al. 2018, Hu et al.

2014, Jayaprakash et al. 2014). For example, Macfadyen & Dawson (2010) presented a proof of concept study to provide weekly predictions of student performance. In this study, a logistic regression model was trained, using a combination of students' VLE interactions and online assessment variables, extracted from fully online courses.

Furthermore, considerable effort has been conducted in the area of developing early warning instruments to predict students who are at academic risk early in the semester on an institutional level. Various data types, LA approaches and prediction techniques have been used to move towards achieving this target. The purpose of these instruments is to allow for early identification and deliver meaningful support for students in need, which helps to optimise their learning outcomes. Although the main focus is on utilising parameters gained from online learning activities in developing such systems, others were interested in additional information driven from other sources. A range of popular prediction approaches were studied and compared by (Howard et al. 2018, Hu et al. 2014, Jayaprakash et al. 2014) to develop early warning systems.

For instance, in the early warning system developed by Hu et al. (2014), the authors experimented with and compared the predictive power of three prediction techniques, using data about students' interactions with VLE components alone. The compared approaches included a decision tree (C4.5), classification and regression tree (CART), and logistic regression. Evaluation of the results showed that the CART approach, accompanied by an adaptive boosting algorithm, produced the most accurate predictive results, where the nominated model was used to serve as a predictive instrument in the developed system.

In other research, Jayaprakash et al. (2014), presented an early alert system to predict academic risk. The prediction model fed in online behaviour data extracted from a collaboration and learning environment, alongside students' demographic factors and past performance parameters. The captured data were fed into logistic regression, support vector machine, naive bayes and decision trees models to compare the performance of each approach. Logistic regression showed the highest power of prediction among the evaluated models.

Moreover, Howard et al. (2018) proposed an early warning system to forecast students' final grades, which enabled identification of students at-risk of failing the course. They examined the predictive power of bayesian additive regressive trees, random forest, neural network, k-nearest neighbours and XGBoost alongside several regression approaches. The researchers applied students' VLE interactions data combined with demographic and internal assessment information for each approach. The results indicated that a bayesian additive regressive trees approach had the best performance.

Villagra-Arnedo et al. (2017) utilised an alternative design for an early prediction system to predict academic performance in terms of a prediction outcomes format and development methodology. The authors classified the students' performance into a three-level classification schema: high, medium and low performance, based on their grades. Ten independent support vector machine models were built to predict the students' performance, where each model corresponded to a prediction week in the semester. The prediction models were fed with data captured from VLE activities and online assessment factors.

Due to the valuable benefits of early warning systems in educational contexts, several international higher education institutions have recruited such systems to provide timely identification and support for risky behaviours over the last decade. Bin Mat et al. (2013) and Sclater et al. (2016) reviewed a range of institutional attempts to develop and use early predictive systems of students' performance in Australia, Germany, the UK and the US.

Moreover, multiple studies have proposed actionable plans to use with students who are at-risk (Choi et al. 2018, Na & Tasir 2017, Wise 2014, Wong & Li 2018) and highlighted the best period in which to support the students (Conijn et al. 2017, Howard et al. 2018). For instance, Jayaprakash et al. (2014) proposed two intervention strategies to deal with students who were identified as being at-risk, where those students were subjected either to receiving a general awareness message or a message that encouraged the learner to join an online academic support environment. Hu et al. (2014) proposed an intervention strategy in their work as follows: when students are detected as being at-risk, the lecturers and students are notified via email and through VLE interface. Then, the lecturer interacts with the at-risk students by scheduling a series of face-to-face tutoring and consultation appointments, as needed. Howard et al. (2018) identified week 5/6 of a 12-week semester as being the critical period to forecast students' likelihoods of success or failure within a course, given that, at this period of the semester, the proposed prediction model starts to produce reasonably accurate estimates. Also, Conijn et al. (2017) indicated after week 3 as being the best time for early intervention, as the model starts to provide accurate classification results at that point.

Utilising LA-powered tools to deliver proactive interventions to in-need students has impacted positively on improving university students' academic achievements and increasing retention rates. Several attempts to identify and support struggling college students reported improvements in academic outcomes (Arnold & Pistilli 2012, Larrabee Snderlund et al. 2018, Sclater et al. 2016, Wong 2017). Larrabee Snderlund et al. (2018) and Sclater et al. (2016) highlight a number of case studies that evaluate the effectiveness of early academic interventions on students' achievements in higher education contexts. For example, the chosen early warning system at Purdue University leads to an improvement in students who obtain B and C grades by 12 percent and the number of students who passed with lower grades declined by 14 percent. Furthermore, Sclater et al. (2016) reported another example at New England University where a trial of LA tools reduced the attrition rate from 18 to 12 percent.

In another comprehensive review of the impact of academic interventions on students' success, Wong (2017) stated that providing academic interventions improved students' outcomes over non-intervention groups of students in all the higher education institutions reviewed. However, the author reported that the effect size of the interventions was varied across the institutions.

Furthermore, Cassells (2018) reported an improvement in students' grades of eight percent on average as result of facilitating timely interventions for students who were flagged as at-risk. Furthermore, Jayaprakash et al. (2014) studied the contribution of timely interventions on groups of students who had been labelled as at academic risk. The supplied academic interventions enhanced the students' grades over nonintervention sets of students by six percent. A recent study reported a seven percent increase in success rates following the implementation of such an intervention strategy. However, other studies have recorded the limited influence of early academic support on students' academic performance (Dodge et al. 2015).

Moreover, multiple studies have reported the positive impact of early interventions on students' completion rates. Arnold & Pistilli (2012) observed a significant improvement in completion rates, by up to 25 percent, over groups of students who had not received intervention support, while Milliron et al. (2014) reported that early intervention increased retention rates by three percent. In a Brazilian University, researchers found that employing an intervention approach resulted in a decline in the drop-out rate of 11 percent (Cambruzzi et al. 2015). However, Jayaprakash et al. (2014) observed an 11.5 percent increase in the probability of withdrawal for students on the intervention list, when compared with their peers.

There are several well-known examples of utilising early warning systems to flag struggling students in practice. One of the first early systems in practice was Course Signals. Course Signals (CS) is an LA-powered early warning and intervention solution to enhance students' success (Arnold & Pistilli 2012). In 2007, Purdue University presented the Course Signal (CS) system. Then, in 2009, it launched an automated version of the system. The CS developers applied various types of student variables to a predictive student success algorithm to compute the students' probability of success. The student information consisted of a combination of demographic and past academic performance information, alongside dynamic data extracted from VLE. The system was integrated with the Blackboard VLE to supply real-time indicators to students about their performance and then provide feedback. The colour of a traffic light signal (red, yellow or green) is presented on the student's personal course interface to indicate their risk level.

Moreover, there are numerous other successful examples of utilising early warning instruments to improve students' academic performance in practice (Sclater et al. 2016). For example, Rio Salado College, in the USA, developed the Rio-PACE model to evaluate students' engagement and progress in a fully online course and, consequently, report students who were at-risk (Smith et al. 2012). In another case, the NTU Student Dashboard developed at Nottingham Trent University, in the UK, is one of the most prominent LA projects to improve college student retention and other objectives (Sclater et al. 2016). All students enrolled at the university are affected by the NTU Student Dashboard and students in-need receive direct assistance. Another initiative established by Edith Cowan University, Australia, implemented the Connect for Success (C4S) program (Jackson 2012). C4S aims to improve students' success by identifying and supporting students who are in need of help. Also, a NYIT STAR model was utilised at the New York Institute of Technology, in the USA, to boost students' retention rates through early interventions for freshman students who had been identified as at attrition risk (Agnihotri & Ott 2014). Furthermore, Bin Mat et al. (2013) reviewed other projects used as early alert instruments to support students such as E^2 Coach used by the University of Michigan, and the Individual Learning Plan (ILP) system used by Sinclair College.

Finally, various higher education institutions have invested in developing predictive instruments for student success, due to their valuable benefits in improving academic outcomes. Several studies have compared a wider range of popular prediction algorithms for identifying the most accurate approach, then utilised the results to develop early warning systems, whilst regression algorithms are the most frequently employed approaches. In addition, many studies have observed the positive impact of utilising early warning systems to deploy actionable interventions on students' success, which help to improve academic outcomes. Several successful institutional examples have been presented that early academic interventions lead to improvements in students' grades (Cassells 2018, Dodge et al. 2015, Sclater et al. 2016) and increases in retention rates (Arnold & Pistilli 2012, Cambruzzi et al. 2015, Milliron et al. 2014). Therefore, the promising results may motivate higher education instantiations to employ such instruments in order to gain the highest benefit from them.

2.4 Summary

During the last decade, much research has been conducted to identify students' personal, academic and learning characteristics, which are then correlated with academic outcomes, and employed to forecast course achievements. Table 2.1 highlights various studies conducted between 2012 and 2018 to develop automated predictive models of academic performance in blended and fully online settings. While most of the reviewed predictive models target academic performance in the form of final achievements, other models are concerned with identifying potential students at-risk of failure to complete their academic courses successfully.

In terms of data sources, a wide range of variables has been extracted from various sources to identify characteristics associated with academic achievement, which have been used to build predictive models. It is common practice to integrate different types of features to feed predictive student success models. Dynamic data are the most popular sources of features, so most predictive models rely partly or solely on this type of data. However, dynamic-related features generally result from purely quantitative analyses of interactions with online learning components. A limited number of studies have considered applying social or qualitative analyses of students, such as applying social network analyses and text-based analyses. Similarly, student-related demographic factors are also a popular source of attributes used to train predictive models, alongside academic-related aspects, such as pre-enrolment and post-enrolment performance, interim assessments and attendance records.

A wide range of prediction approaches have been utilised in the literature to forecast students' academic performance in higher education settings. On the other hand, a few researchers have applied statistical, genetic programming or proposed novel approaches to predict academic performance, whilst machine learning algorithms are the most frequently-used approaches to build predictive instruments. The clear majority of the reviewed research has performed a comparative analysis on different collections of predictive techniques, where diverse algorithms were studied; however function-based methods are the most popular approaches, followed by tree-based algorithms. Function-based approaches include variations of techniques such as logistic regression, linear regression and, stepwise, forward and backward regression and use of support vector machine. Furthermore, several tree-based classification algorithms have been examined to build forecasting models including C4.5, CART, ID3, J48 and regression tree algorithms. Other machine learning approaches were analysed and compared in multiple studies including bayesian, ensemble learning, rules-based and clustering algorithms. However, training dataset can have a significant influence on the performance of machine learning algorithms.

However, most of the past effort has focused on training and testing predictive models on populations drawn from one dataset (Cassells 2018, Dominguez et al. 2016, Hayes et al. 2017, Howard et al. 2018, Hu et al. 2014, Jishan et al. 2015, Lopez et al. 2012, Romero et al. 2013), which usually belongs to a single course. Limited studies have used independent datasets collected from different courses or academic periods to evaluate the models. Furthermore, cross-validation is the most commonly-used validation and dataset partitioning method (i.e (Romero et al. 2013, Smith et al. 2012, Wolff et al. 2013, Xing et al. 2015)).

Moreover, the vast majority of the reviewed works built and evaluated data collected from courses taught at one institution, which raises concerns about the scalability of researchers' outcomes where datasets contain the learning, performance and demographic information associated with a particular educational and cultural context. Therefore, the models fed with those datasets might only be useful within similar settings.

This thesis employs various quantitative, qualitative and social approaches to analyse students' VLE activities and contributions to online discussion forums to produce a set of variables. The resultant variables were utilised to develop an exemplar multi-course early warning framework of academic risk in higher education settings that provides early and accurate predictions. The proposed exemplar multi-course early warning framework's predictivity power was examined against unseen course data to simulate real-life scenarios. Moreover, we extended our work to allow the framework to learn from its additional datasets, provided by users. Building an updateable early warning framework fills the gap regarding the scalability of the predictive framework in new educational environments, as well as providing a fully-automated framework to enable continuous improvements to the underlying predictive instrument by incrementally increasing the amount of training dataset information fed into the system.

This chapter has provided a background of the efforts that have been undertaken to develop predictive models of students' performance in respect of the data and approaches used to build the models. It has discussed the different types of student characteristics used to predict academic performance in higher education settings. In addition, this chapter has reviewed various studies aiming to develop predictive models. Furthermore, this chapter has presented concerns from past efforts about developing early warning systems for academic success and, in some cases, of adopting such systems in practice. Finally, the chapter concluded by summarising the efforts made in developing predictive models of students' final achievements and then has drawn attention to the gaps in the research field.

The next chapter describes how this thesis contributes to the field of predicting students' performance in the light of the identified gap. Chapter 3 defines the efforts made throughout this study to bridge the gap in the literature, while Chapters 5, 6 and 7 present detailed descriptions of this work.

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Aguiar et al. (2014)	Static Dynamic Academic	Bayes-based Function-based* Tree-based	Blended	1 course	429	Same Dataset	Y	Attrition Risk
Ashenafi et al. (2015)	Dynamic Assessment Survey	Function-based	Blended	2 courses	206	Independent Dataset		Final Performance
Ashenafi et al. (2016)	Dynamic Assessment Survey	Function-based	Blended	2 courses	229	Same Dataset		Final Performance
Azcona & Casey (2015)	Dynamic	Bayes-based Function-based Lazy-based Tree-based*	Blended	1 course	NA	Same Dataset		Academic Risk
Bainbridge et al. (2015)	Static Dynamic Academic Assessment	Function-based	Online	NA	1,073	Same Dataset		Academic Risk

Table 2.1 continued from previous page

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Bayer et al. (2012)	Static Dynamic Academic	Bayes-based Function-based Lazy-based Rule-based* Tree-based	Blended	Multiple courses taught over 3 years	4,373	Same Dataset	Y	Attrition Risk
Bydzovska (2016)	Static Dynamic Academic	Bayes-based Function-based* Rule-based Tree-based	Blended	138 courses taught between 2010-2013	48,031	Independent Datasets		Final Performance
Cen et al. (2016)	Dynamic Academic	Function-based Tree-based*	Blended	1 course	122 students, 72 groups	Same Dataset		Individuals and Groups Performance
Chai & Gibson (2015)	Static Dynamic Academic	Function-based* Tree-based	Blended	Freshman coursers taught between 2011-2013	23,291	Same Dataset	Y	Attrition Risk
Conijn et al. (2017)	Dynamic Assessment	Function-based	Blended	17 courses	4,989	Same Dataset	Y	Final Performance

Table 2.1 continued from previous page

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Dascalu et al. (2016)	Dynamic	Function-based Statistical	Blended	1 course taught over 6 semesters	148 students	Same Dataset	Y	Academic Risk
Howard et al. (2018)	Static Dynamic Assessment	Ensemble-based Function-based Lazy-based Tree-based*	Online	1 course	136	Same Dataset	Y	Academic Risk
Hu et al. (2014)	Dynamic	Ensemble-based* Function-based Tree-based	Online	1 course	300	Same Dataset	Y	Academic Risk
Jayaprakash et al. (2014)	Static Dynamic Academic Assessment	Bayes-based Function-based* Tree-based	Blended	Multiple courses taught over 3 semesters	15,150	Independent Dataset		Academic Risk
Jishan et al. (2015)	Academic Assessment	Bayes-based* Function-based* Tree-based	Blended	1 course enrollments during 18 months	181	Same Dataset		Final Performance

Table 2.1 continued from previous page

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Lopez et al. (2012)	Dynamic	Bayes-based* Classification via Clustering Function-based Rule-based Trees-based	Blended	1 course	114	Same Dataset	Y	Final Performance
Mueen et al. (2016)	Static Dynamic Academic Assessment	Bayes-based* Function-based Tree-based	Blended	2 courses	60	Same Dataset	Y	Academic Risk
Natek & Zwilling (2014)	Static Academic Aassessment	Tree-based	Blended	1 course taught over 3 years	106	Independent Datasets		Final Performance
Osmanbegovic & Suljic (2012)	Static Academic Survey	Bayes-based* Function-based Tree-based	Blended	Multiple courses offered during summer semester	257	Same Dataset	Y	Academic Risk

Table 2.1 continued from previous page

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Pardo et al. (2016)	Dynamic Assessment	Tree-based	Blended	1 course	272	Same Dataset	Y	Midterm and Final Performance
Rogers et al. (2014)	Static Dynamic Academic	Index Method Function-based	Blended	1 course taught over 2 years	2,332	Independent Datasets		Academic Risk
Romero et al. (2013)	Dynamic	Bayes-based Classification via Clustering* Function-based Rule-based Tree-based	Blended	1 freshman course	114	Same Dataset	Y	Final Performance
Shelton et al. (2016)	Static Dynamic	Ensemble-based Function-based Rule-based Tree-based*	Online	12 courses	509	Same Dataset	Y	Academic Risk

Table 2.1 continued from previous page

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Shelton et al. (2017)	Static Dynamic	Ensemble-based Function-based Rule-based Tree-based*	Online	18 courses	661	Same Dataset		Academic Risk
Smith et al. (2012)	Dynamic	Bayes-based	Online	1 course taught over 2 semesters	539	Same Dataset	Y	Academic Risk
Villagra-Arnedo et al. (2017)	Dynamic	Function-based	Blended	1 course	336	Same Dataset	Y	Academic Risk
Wolff et al. (2013)	Static Dynamic Academic Assessment	GUHA Data Mining	Online	3 courses	7,701	Same Dataset	Y	Academic Risk

Table 2.1 continued from previous page

Study	Attributes Sources	Approaches	Learning Setting	Data Source	N	Training Paradigm	CV	Target
Xing et al. (2015)	Dynamic	Bayes-based Genetic Programming* Function-based Rule-based Tree-based	Online	1 course	122	Same Dataset	Y	Academic Risk

Table 2.1: Summary of the literature on performance predictive instruments in higher education. For approaches, * symbol refers to the approach(s) which provide the most accurate predictions in comparison studies. The abbreviation NA means that certain information is not available.

Chapter 3

An Adaptive Multi-Course Early Warning Framework for At-Risk Students

This chapter presents a high-level explanation of the work included in the thesis. It describes how this study is contributing to the research area of identifying at-risk students in higher education contexts, where the strategies described aim to fill the gap identified in the literature. This study tackles various aspects of predicting students' performance, including introducing and evaluating alternative methods to extract academic risk predictors other than those used frequently in the current literature. Furthermore, in this thesis, we have proposed a novel design for a Grey Zone to enhance the quality of binary classifiers. Moreover, this thesis presents an exemplar multi-course early warning framework for at-risk students, alongside proposing and evaluating a dynamic strategy to enhance its adaptivity.

This chapter is organised as follows: Section 3.1 presents briefly VLE and a discussion forum predictor extraction methodological gap, along with approaches proposed to bridge it. A proposed extraction method for diverse variables has been utilised and evaluated in this work to predict students who are at-risk academically. Then, Section 3.2 provides an overview regarding the novel design of the Grey Zone proposed in this study to enhance the performance of binary classifiers. Section 3.3 outlines the gap in the literature in terms of developing an early warning framework of at-risk students that can be used to predict instances drawn from multiple course contexts. Section 3.4 briefly describes the efforts made towards addressing the scalability and updatability issues identified in the literature when developing student performance prediction instruments. The solution involves proposing an adaptive strategy that allows the prediction instruments to cope dynamically with any changes which may arise in the prediction space.

3.1 VLE and Discussion Forum Variables

A major vehicle for accurate detection of at-risk students is utilising quality predictors. Online activities data provide a valuable source of information that can be used to identify early signs of students' academic risk within the underlying course context. However, in a blended learning model, relying solely on such data is a challenging task as this learning mode is designed to hybrid off-line and online learning activities in conjunction with each other.

In fact, students' off-line learning activities are usually not recorded, which makes it hard to monitor such learning patterns. Therefore, it is vital to explore different aspects of online learning data to identify the most influential online risk characteristics. While most of the previous work relies on quantitative analysis of such data (i.e. (Cen et al. 2016, Chai & Gibson 2015, Conijn et al. 2017, Mueen et al. 2016, Pardo et al. 2016, Villagra-Arnedo et al. 2017)), a few studies performed qualitative analysis to forecast students' achievements (i.e. (Crossley et al. 2016, Haig et al. 2013, Lopez et al. 2012, Robinson et al. 2016)). Qualitative analysis offers an opportunity to open new doors to explore hidden elements of students' learning and social experiences.

This thesis addresses a literature gap regarding qualitative extraction methods of students' characteristics, by performing qualitative analysis approaches on different types of online engagement data. Qualitative approaches involve methods to measure the change in students' engagement patterns over time regarding various aspects of learning and mining textual content created by students, alongside other quantitative methods.

In this work, we introduce an automated language-based analysis mechanism to investigate the content of students' messages posted on virtual discussion forums of diverse courses. The approach works by scoring the strength of students' sentiments expressed in discussion forum posts, based on corresponding adverbs. To perform the adverb weighting task robotically, a Digital Adverb Strength dictionary has been developed, which contains 3,762 English adverbs currently in use. A detailed description of the Digital Adverb Strength Dictionary development process is outlined in Section 4.3.

Moreover, the influence of the extracted predictors using the proposed qualitative methods were ranked and compared against other predictors extracted from discussion forum participation data in thirteen blended learning courses. A ranking method was performed using multiple well-known feature selection methods.

3.2 Grey Zone Design

In a binary classification context, classifiers are traditionally designed to follow black-and-white decision-making strategies to determine the prediction class. Identifying at-risk students' binary classification instruments is no exception (i.e. (Azcona & Casey 2015, Bainbridge et al. 2015, Bayer et al. 2012, Chai & Gibson 2015, Dascalu et al. 2016, Hu et al. 2014)). The chosen decision-making strategy forces binary classifiers to make marginal decisions regarding prediction classes, which limits their ability to distinguish actual instance classes when underlying features reflect similar observations for instances belonging to different prediction classes.

Therefore, to enhance the ability of binary prediction instruments to produce quality outcomes, we proposed a novel concept, called Grey Zone design, where prediction probabilities are divided into three different zones. The design works by generating a new zone that overlaps both classes in the prediction space where instances fall in the Grey Zone. They are then subject to further investigation using alternative classifiers. The design suggested that Grey Zone boundaries are identified by defining the weakness of a base predictive model, where it fails to provide quality predictions by performing error analysis metrics. The base model is responsible for producing initial prediction decisions by referring instances under prediction to one of three zones: white, black or grey. Then, instances that fall in the Grey Zone are subject to further prediction, using specially implemented Grey Zone models that are expert in distinguishing prediction classes in such circumstances.

In this thesis, we examined the effect of the Grey Zone concept on the predictive performance of a binary classification model in terms of forecasting students' performance. A set of experiments were conducted to investigate the impact of the proposed Grey Zone design on model's performance, comparing this with traditionally-used decision-making strategies on a weekly basis. Each experiment is run to evaluate and compare the quality of two decision-making strategies with each other. Furthermore, the Grey Zone design has been utilised to develop the multi-course early warning framework for at-risk students proposed in this study.

3.3 A Multi-Course Early Warning Framework for At-Risk Students in Blended Learning Setting

Developing reliable early warning instruments to detect potential at-risk students is a critical step in order to deliver proactive and timely interventions for those students. In the field of predicting tertiary students' performance, the main focus is on developing prediction instruments using data collected from a single course (Cassells

2018, Dominguez et al. 2016, Hayes et al. 2017, Howard et al. 2018, Hu et al. 2014, Jishan et al. 2015, Lopez et al. 2012, Romero et al. 2013). However, in higher educational contexts, courses are varied in terms of course structure, required workload and assessments. These facts raise concerns about the applicability of these prediction instruments on instances drawn from other courses.

Moreover, it is widely assumed that courses have homogeneous data distribution. Hence, researchers employ cross-validation approaches to build and validate prediction instruments (Romero et al. 2013, Smith et al. 2012, Wolff et al. 2013, Xing et al. 2015). A limited amount of research has utilised independent validation dataset in the development process or employed an unseen validation dataset such as in (Bydzovska 2016, Chai & Gibson 2015, Rogers et al. 2014).

The proposed exemplar multi-course early warning framework for at-risk students takes into consideration the issues in the existing literature. The proposed framework has been developed using data drawn precisely from students' VLE interactions and discussion forum participation data. It aims to provide early, quality predictions of students who are at academic risk across multiple Computer Science courses that follow blended learning pedagogies. The prediction instrument has been designed using the proposed novel Grey Zone concept described in this study.

The development of a multi-course early warning framework bridges a methodological gap by developing a prediction instrument that is able to provide early, quality predictions across multiple courses, whilst also anticipating the framework's performance in future courses. Therefore, the proposed framework is evaluated using an unseen dataset, drawn from multiple heterogeneous courses offered in different academic periods, where every course has its own unique structure and online activities distribution. Moreover, by analysing the evaluation results, it is possible to identify the optimal week in which course coordinators and lecturers should start intervening. The identified intervention timing is the earliest week where the framework starts to produce quality predictions across all evaluated courses. Pointing to the earliest intervention week helps to determine how early the predictive framework can provide reliable, high-quality outcomes.

3.4 Adaptive Strategy

When predicting students' academic achievements in the literature, the majority of the extant prediction instruments are devoted to static machine learning environments, where models tend to be trained on historical information and remain fixed due to the absence of an adaptive mechanism (Bainbridge et al. 2015, Cen et al. 2016, Chai & Gibson 2015, Howard et al. 2018, Hu et al. 2014, Jishan et al. 2015, Natek & Zwilling 2014). Moreover, these prediction models are built and evaluated using data collected from historical courses taught at one institution, which raises concerns about the scalability of the researchers' outcomes, where in-suite datasets contain learning, performance and demographic characteristics associated with a particular educational and cultural context. Furthermore, these concerns extend to include the predictive models' capacity to cope with changes which may occur in students' learning styles, over time. Utilising an adaptive mechanism allows predictive instruments to adapt to new cultural or educational settings, alongside updating their properties dynamically to maintain and enhance their prediction quality.

On the other hand, updating predictive models using additional courses data is typically possible at the end of academic semester when results become available. An alternative approach can be utilised is incrementally updating a predictive model during the semester whenever a student withdraws for the course (Lagus et al. 2018). In this approach data belong to drop-out is fed the predictive model.

There is a gap in the literature regarding the absence of dynamic adaptive strategies that enhance the scalability and updatability of student performance prediction instruments. To fill this gap, we developed an adaptive framework that allows an underlying prediction model to cope with changes in the prediction space, alongside adapting to new prediction environments. To the best of our knowledge, this is the first work introducing adaptive learning concept in this research context.

This study introduces the Adaptive Grey Zone Model (AGZEM) framework, which supports predictive instrument to adapting to changes in students' learning behaviours. The developed framework is aligned with the novel Grey Zone strategy proposed in this work. The development of an AGZEM framework involves also building two complementary adaptive algorithms to update the predictive components, as well as Grey Zone configurations. The first algorithm is designed to update the base and Grey Zone predictive models, while the second computes the optimal Grey Zone boundaries. The focus was on developing adaptive mechanisms that suit a multi-course early warning framework, to cope with any changes that may occur in the prediction environment. Therefore, we take into the consideration the design of the prediction models used in the multi-course framework, while designing the new, adaptive strategy.

The feasibility of the proposed adaptive framework and algorithms was tested experimentally over two adaptive scenarios. The evaluation process was deployed with and without historical data forgetting criteria to examine adaption outcomes over both scenarios.

The next chapter presents the data gathering and preparation processes, as well as a description of the extracted variables, tools and methods used in this work. Chapters 5, 6 and 7 present the implementation and evaluation of the work involved in this thesis in order to develop an adaptive multi-course early framework to detect at-risk students in Computer Science courses. The multi-course framework that was developed is featured to provide early, quality predictions, alongside the ability to learn robotically from any additional data batches that may obtained in the future.

Chapter 4

Context and Data

This thesis investigates students' online learning behaviours and contributions data in blended learning setting to detect students who are potentially at academic risk early in the semester. Quantitative, qualitative, and social analysis approaches were performed on students' data drawn from VLE to compute a range of student-level characteristics. The extracted student characteristics were analysed to determine significant predictors that reflected each student's academic risk status. Then, influential predictors were used to build a computational multi-course early warning framework that forecasts students who are at-risk in a timely manner. Furthermore, this study involves developing adaptive mechanisms, which allow the framework to update its properties to cope with changes in online learning patterns.

This chapter describes the data, methods and tools used in this thesis. It is organised as follows: while Section 4.1 discusses the collected online learning data and the participants involved in the study in detail, Section 4.2 presents the data preparation methodology and describes the features extracted from the gathered data. Section 4.3 presents the development process for the Digital Adverb Strength dictionary implemented for this study. Section 4.4 provides a background to the CoreNLP toolkit that was used for the language-based analysis of the students' generated texts. Then, Section 4.5 presents machine learning approaches that was utilised to construct the predictive instruments used to determine those students at academic risk. Section

4.6 discusses the evaluation metrics used to ensure the quality and accuracy of the developed predictive models. Finally, Section 4.7 presents a summary of the chapter.

4.1 Data Collection

This study relies solely on features calculated from students' VLE interactions and online discussion forum participation data. The obtained data can reflect the student's degree of engagement with the online learning components, sentiments expressed in course-related posts and social presence in an online learning community. The student data were drawn from undergraduate and postgraduate Computer Science courses taught at the University of Adelaide in Australia. The participating courses were offered in both the first and second semesters, over multiple academic periods.

The University of Adelaide's academic year consists of two main semesters, where each semester consists of 12 core teaching weeks. One optional teaching week follows the core study period. In each semester, there is a 2-week mid-semester break, which typically falls between weeks 5 and 8 of the core study weeks. All of the gathered courses were designed to follow blended learning theory and were mediated by a Moodle VLE. The rest of this section describes the dataset collected, alongside defining the participants and explaining how their privacy was maintained. In addition, it presents a statistical analysis of the collected datasets.

4.1.1 Dataset Structure

Data were collected from two different online sources. The first source contained the students' VLE interactions and participation data that had been retrieved from the Moodle VLE database. Moodle is a well-known VLE that has been used in many higher education institutes across many years (*Moodle* 2017). The two online learning datasets associated with each course are the Moodle logs dataset and the discussion forum participation dataset. The logs dataset contains records of students' VLE actions, which were retrieved to CSV format files. The other dataset was drawn in an XML format, and includes discussion forum data. The second source consists of the students' final performance records, which were retrieved from the university database in CSV format files. Students who choose not to complete their courses and dropped out before the deadline were tagged as dropping out, with no awarded mark.

The VLE interaction logs dataset includes detailed data of students' actions performed on the Moodle VLE. Within the dataset, each row presents a record of a single activity that covers six entries, which are:

- The Date and Time at which the activity event occurred
- The User ID that the user used to access the system
- The component which the user accessed
- The event name and the type activity performed
- A description of the object which was affected by the activity.

There are a range of activities that can be performed on the Moodle VLE, based on the user's role. Moodle allows users to access the system as administrators, course creators, lecturers, students or guest users. Hereafter, we will cover major student-level activities that are relevant to this study. There are two sets of keywords used to identify the types of action performed on a Moodle VLE. The first version represents the set of keywords that appear in courses taught before 2016, while courses taught during and after 2016 use the second version of the keywords to denote the exact same students' actions on Moodle VLE.

- Course Access
  - *Course View (Version 1 and Version 2) denotes:*
    User viewed course module on the VLE.
- Discussion Forum
  - *Forum Add Discussion (Version 1) and Version 2: Discussion Created denote:*
    User created a new thread in the course discussion forum.
  - *Forum Add Post (Version 1) and Version 2 Post Created denote:*
    User added a new post that replies to a thread or post.
  - *Forum Update Post (Version 1) and Version 2 Post Updated denote:*
    User updated an existing post.
  - *Forum View Forums (Version 1) denote:*
    User viewed all discussion forums available.
  - *Forum View Forum (Version 1) and Course Module Viewed with 'forum' in the description (Version 2) denote:*
    User viewed all the discussion titles available on the forum.
  - *Forum View Discussion (Version 1 and Version2) denote:*
    User viewed a discussion on the forum.
- Resources Components
  - *Resources View (Version 1) and Course Module Viewed with 'resource' in the description (Version 2) denote:*
    User viewed a resource that has been uploaded to the course module.
- Assignments Components
  - *Assignment View (Version 1) and Course Module Viewed (Version 2) denote:*
    User viewed an assignment description that has been uploaded to the course module.

The other dataset involves discussion forum contents. This dataset was retrieved in an XML format, which carries data about forum structures and messages generated by users. XML uses tags to carry discussion forum data, including the discussion id, post id, user id, post parent id and the post's subject and contents. In addition, there are tags carrying the time at which the post was created and modified, using the UNIX timestamp.

4.1.2 Data Anonymisation

Identifying individual student information is central to this study. Therefore, we must be able identify the performer of each record. However, maintaining the privacy of the participants is essential to this study. At the data collection stage, the online activities data was anonymised by identifying students using the Moodle VLE subscription id number instead of their actual names or university id numbers. Identical identification numbers are used across the collected datasets, however, there were traces of users' information in the body of some forum posts, particularly in the greetings and signatures sections. Thus, we used the CoreNLP tool, described in Section 4.4, to remove any personal identification information from the textual content as soon as possible after the data was collected.

Furthermore, to minimise the risk of identifying students, all the participants' data are encoded using a combination of "STD"letters and five-digit unique random numbers. The process of replacing the original participant identification numbers with the new ones was done automatically by an application built specially for this purpose, while initial student IDs were not recorded.

4.1.3 Data Description

The data used in this study was collected from thirteen blended Computer Science courses taught at the University of Adelaide, Australia, in the first and second semesters between 2012-2016, and offered to undergraduate and postgraduate students. Data about students' online engagements were retrieved from the Moodle VLE. The sample courses were generally offered for undergraduate students in the form of eight undergraduate-only courses and the remaining five courses were offered for both undergraduate and postgraduate students. Of the thirteen courses, seven courses were taught in 2012 and five courses taught in 2013, while only one course was taught in 2016. Table 4.1 presents an overview of the course names and levels, and the study periods that were used in this study.

Course ID	**Name**	**Level**	**Study Period**
Course 1	Computer Architecture	UG	Semester 1, 2012
Course 2	Artificial Intelligence	UG+PG	Semester 1, 2012
Course 3	Programming Techniques	UG+PG	Semester 1, 2012
Course 4	Computer Graphics	UG	Semester 1, 2012
Course 5	Distributed Systems	UG+PG	Semester 2, 2012
Course 6	Software Engineering and Project	UG+PG	Semester 2, 2012
Course 7	Operating Systems	UG	Semester 2, 2012
Course 8	Advanced Algorithms	UG+PG	Semester 1, 2013
Course 9	Artificial Intelligence	UG	Semester 1, 2013
Course 10	Advanced Programming Paradigms	UG	Semester 2, 2013
Course 11	Algorithms and Data Structures Analysis	UG	Semester 2, 2013
Course 12	Computer Systems	UG	Semester 2, 2013
Course 13	Parallel and Distributed Computing	UG	Semester 1, 2016

Table 4.1: An overview of the courses used in this study. For the Level, the following abbreviations are used: (UG) refers to undergraduate, (UG+PG) refers to undergraduate and postgraduate.

A total of 1,476 enrolments were obtained from the thirteen courses with enrolment rates ranging from 76 to 153 students per course M = 113.5 (SD = 29.2). As several courses were offered in the same or adjacent academic periods, there were 743 unique students enrolled in the thirteen courses used for this study. Most students (402) only appeared in a single course of the collected courses; 147 students enrolled in two courses; 91 students enrolled in three courses; 44 students in four courses; 35 students in five courses; 14 students enrolled in six courses and 8 and 2 students enrolled in seven and eight courses, respectively.

The participating courses are structured to combine on-campus activities and online media in the teaching process. Most courses offered learning resources and assignments specifications, alongside other course-related digital materials uploaded on the VLE, where students have access to them online. Moreover, all courses utilised online discussion forums as communication and collaboration tools. The initial VLE activities datasets of the thirteen courses contain 373,197 logs in total, ranging from 7,199 to 84,881 logs per course M = 28,707.5 (SD = 19,823.5), while the initial discussion forum datasets contain 3,797 posts, the data ranges from 69 to 944 posts per course in total, where M = 292.1 (SD = 222.2). However, lecturers' and other users' VLE activities and posts are included in the initial logs and discussion forum datasets. Figure 4.1 presents a bar chart showing the frequencies of the major activities per course, as they are recorded in the initial VLE log files.

However, course structures and the amount of uploaded resources on VLE are varied per course, as a consequence of students' usage patterns, which are then reflected in the amount of VLE interactions entries. For instance, the software engineering and project course (course 6) is a group project course with individual assessment components, which means it is sufficient for one team member to check the VLE and update the rest of the team. Therefore, the course's nature resulted in a relatively low volume of interactions within the system, when compared with the number of enrolments. In another example, the advanced algorithms course (course 8) has a very limited number of learning resources made available on the VLE, while the algorithms and

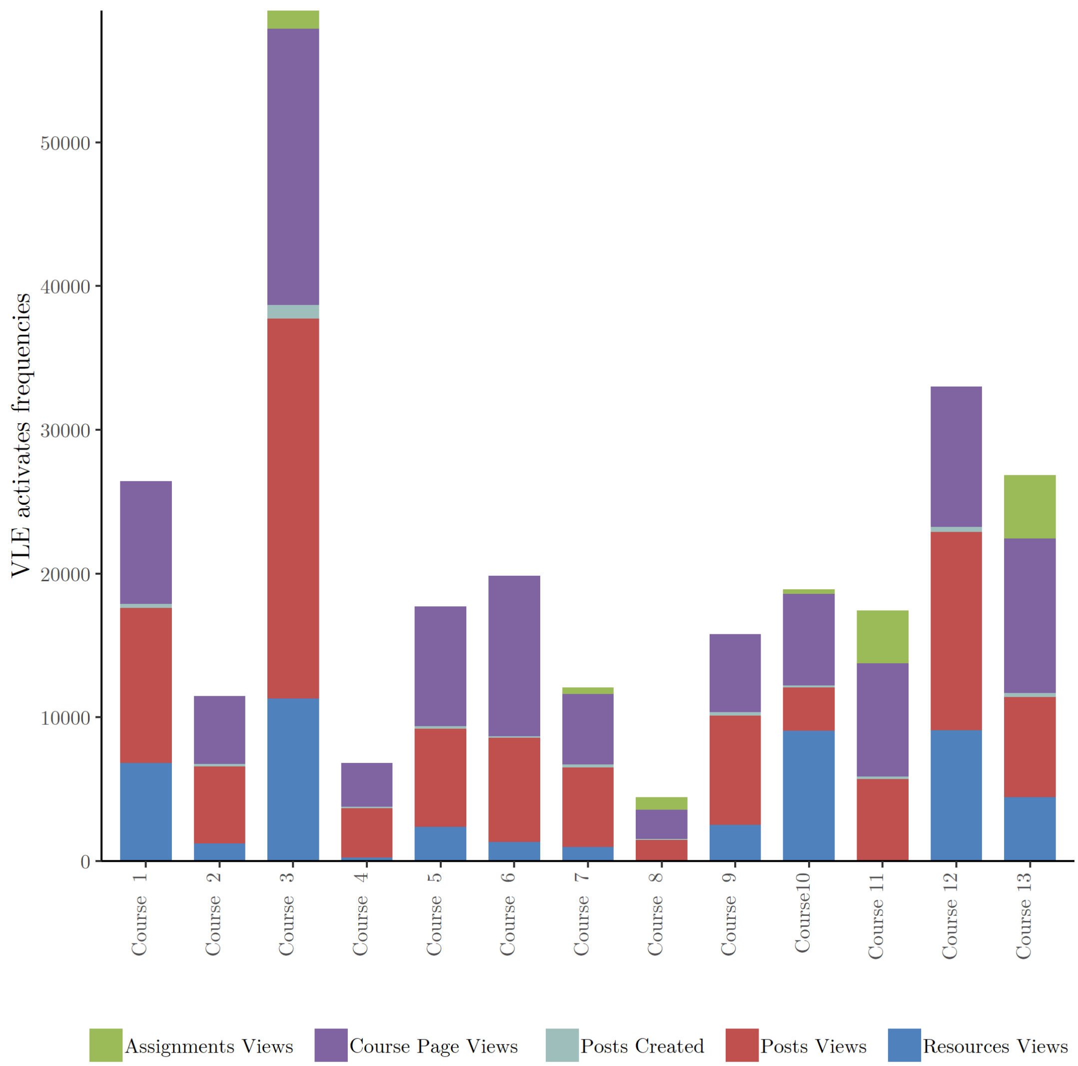


Figure 4.1: The frequencies of activities in the initial VLE log datasets.

data structures analysis course (course 11) has no online learning resources available on the VLE module at all, which explains the limited number or absence of resources view entries in the courses' VLE log datasets. In another case, the programming techniques course (course 3) has very high volume of viewing post actions and interactions recorded with VLE components due to the extensive volume of participation in course discussion forums and the significant amount of resources available on the VLE platform.

Although final course marks are awarded for students who continue to be enrolled in the course, students who withdrew from their courses are labelled as "dropouts", without receiving any final grades. The nature of the assessments and their weightings are diverse, and organised per course. Furthermore, the assessment component may involve various off-line and online sets of tasks such as a final exam, assignments, quizzes, group tasks and programming projects. The majority of the courses investigated in this project used an independent web submission system, offered by School of Computer Science to handle online student submissions, rather than utilising the Moodle VLE which logs submission actions.

The final marks range from 0 to 100 marks, where the minimum mark to pass any course successfully is 50 and any mark less than 50 implies a student has failed the course. A statistical analysis of the participants final performance shows that 52.5 percent of the enrolled students passed their courses successfully, M = 59.6 (SD = 21.8), while about 19 percent received fail grades, M = 21.6 (SD = 8.4). 28.5 percent of participants decided to withdraw from their courses, M = 32.3 (SD = 18.6). In one course, the artificial intelligence course (course 2), there is no record of any drop-out students, as all the students received final marks. On the other hand, some courses have a noticeably higher drop-out rate than others, notably the computer architecture course (course 1) and the algorithms and data structures analysis course (course 11). This can be explained by the fact that students can enrol in multiple courses offered in the same semester, then choose between the course streams that satisfy their preferences later on/in the weeks up to census week. Therefore, a significant a number of students may decide to drop-out. For instance, an analysis of students' enrolment behaviours in the gathered dataset shows that, for courses 1 and 11, about 39 percent of the students who drop-out of these courses have enroled in a number of courses in same semester and choose to withdraw from these particular courses once the semester has commenced. The students' final marks are grouped into five grade categories, as described in Table 4.2. An illustration of the students' final outcome distributions across all courses can be found in Figure 4.2.

Grade	**Description**
High Distinction	Final mark of or over 85
Distinction	Final mark of or between 75 and 84
Credit	Final mark of or between 65 and 74
Pass	Final mark of or between 50 and 64
Fail	Final mark below 50

Table 4.2: Awarded final grades categories.

4.2 Data Preparation

After the data collection stage, it was critical to prepare and organise datasets properly prior to data analysis to avoid any future errors. This study uses ex-post facto students' VLE interactions and discussion forum data, collected from courses taught during various academic semesters. The initial phase was to parse the discussion forums' XML files into CSV files, so as to apply data cleaning and preprocessing procedures. The rest of this section outlines the data preparation process, including dataset cleaning, time-series clustering, labelling dependent variables and data preprocessing methods.

4.2.1 Dataset cleaning

Data cleaning is an essential part of preparing data. It involves the process of detecting and removing corrupted and irrelevant data points. The courses' datasets were cleaned and validated individually, based on their specifications, such as semester and year offering. The original VLE interactions and discussion forum datasets contain records for any user who accessed the course module, regardless of whether or not they are enrolled students. Records of irrelevant users were excluded from the VLE

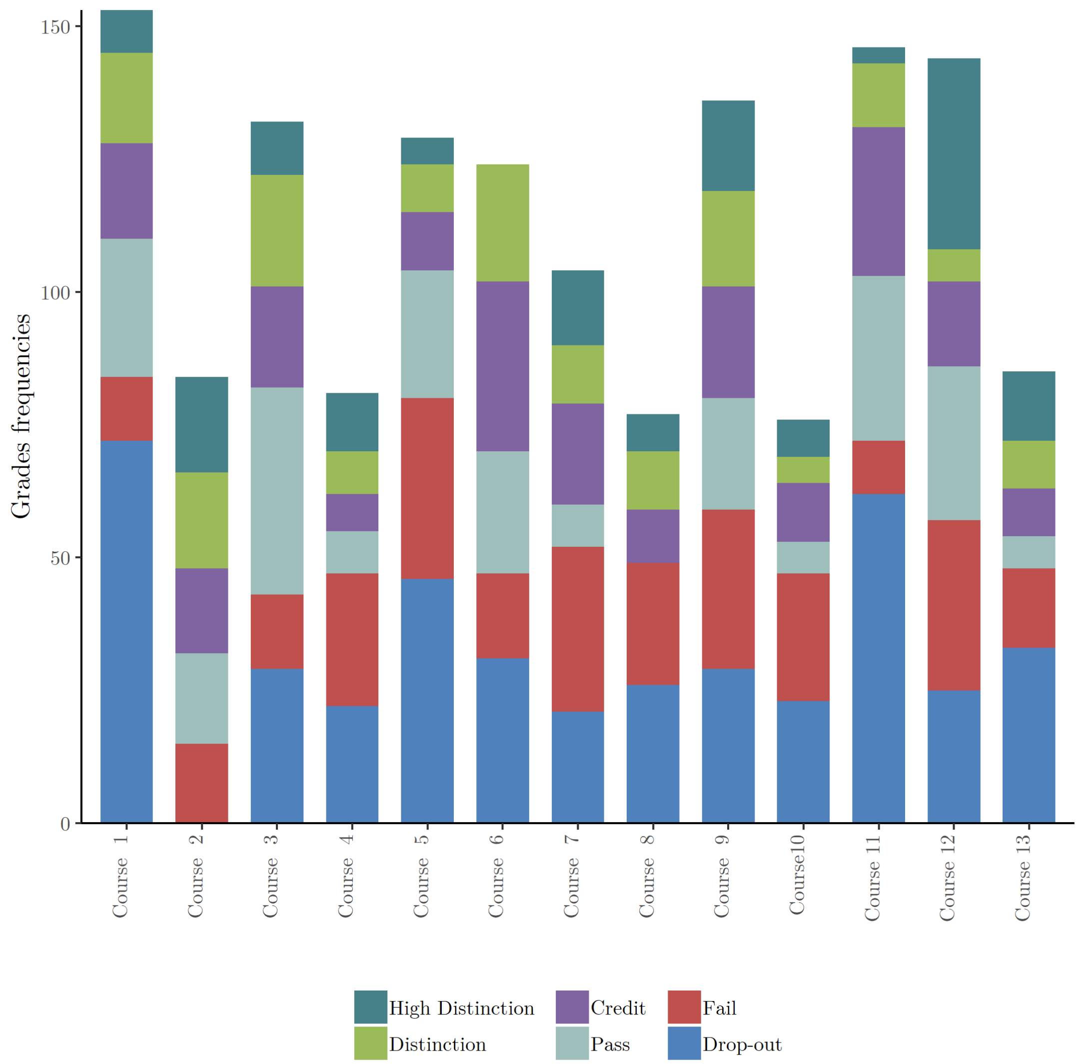


Figure 4.2: Illustration of students' final grades distribution across all courses.

log datasets. However, these cleaning criteria were not applied to the discussion forum datasets, to maintain the social ties with other users.

Furthermore, as most of courses modules are created before the actual semester begins and they are still accessible after the semester ends, the initial datasets contain actions occurring outside each semester's official academic schedule, as used by the university. Those VLE activities that occurred either before the first official core teaching week of the semester or after the end of the core study period were omitted from the dataset. Each field in the dataset was evaluated carefully to detect missing values and corrupted records. Corrupted data fields were either treated by filling in missing values or removing whole records from the course datasets. The data cleaning process resulted in omitting 99,214 superfluous records from the original datasets, M = 7,631.8 (SD = 5,685.4). Therefore, the initial VLE activities datasets were narrowed to 273,983 clean logs, M = 21,075.6 (SD = 15,140.6). Only data cleaning criteria related to post creation times were applied to discussion forum content datasets. This triggered the deletion of 585 posts from the initial datasets, M = 45 (SD = 43.1), which reduced the total number of posts to 3,211 M = 247 (SD = 209.2).

Moreover, this study utilises textual-based analysis on the student-generated texts posted on the courses' discussion forums. Therefore, close attention was to paid to cleaning posted text messages before conducting the analysis. The posts' contents were cleared of unwanted, non-ASCII characters and noisy texts. However, the meanings expressed in posts were maintained and were not affected by the cleaning process. For example, generally posts in course discussion forums contain HTML tags and parts of programming codes, which is considered as noise and extra information. The posts' textual-based cleaning process did not result in omitting any posts from the original datasets.

4.2.2 Time Series Clustering

Time series generation is an important method to measure and use to predict student performance in a timely manner. The students' VLE and discussion board activities were grouped into weekly patterns for 14 weeks, which consisted of the 12 core teaching weeks and the 2 weeks of the mid-semester break. Since the mid-semester break weeks differed each semester, these weeks' data were labelled as "Break-week". The time series clustering method relies on the actual academic calendar dates to determine the start and end weeks in each semester, in addition to the mid-semester break weeks, whilst Sunday is considered to be the first day of the week. Subsequently, data is stored in datasets (DS) that are defined as follows:

VLE_activities_DS = {*Date and Time, User ID, Component, Event Name, Description, Event context*}
Forum_post_DS = {*Date and Time, User ID, Receiver, Post ID, Discussion ID, Message*}

4.2.3 Independent Variables

Students' activities within VLE and online discussion forums reflect their learning and engagement level in the course. This thesis relies on data generated by students' engagement with the VLE objects and virtual forum contributions as the sources of independent variables. Students' participations records were analysed to calculate the desired variables on a weekly basis. A collection of 53 variables were extracted to evaluate their relationship with each student's academic risk status. This section presents those independent variables computed, based on various types of analytical approaches.

VLE interactions Variables

In blended learning courses, VLEs are used to support the learning process. Each course module accommodates different VLE elements, in terms of the nature and number of learning tasks, activities, and available resources. A number of major VLE elements are frequently utilised across the majority of each course's VLE modules. This study employs the commonly performed VLE actions as a source of independent variables, which have been used across the collection of courses. An overview of the original and derivative independent variables extracted from the VLE logs are to be found in Table 4.3 where Quantitative, Qualitative and Social variable's types refer to analytical approach used to compute underlying variable while binary type refers to nature of data held in the variable.

Variable	**Description**	**Type**
Appended course page view frequency	Total number of times a student viewed the course module from week 1 up to the current week	Quantitative
Course page view frequency	Total number of times a student viewed course module in the current week	Quantitative
Course page views trend	The trend of course module view pattern over time; data sequences is calculated on a weekly basis starting from week 1	Qualitative
Course page views trend flag	A flag to denote if a computed course module view trend value is positive	Binary
Change in course page view frequency over 1-week	The change in viewing course module frequencies between current and prior week; if the specified week is 1, this variable will be equal to the week 1 course module views' frequency	Quantitative
Change in course page views over 1-week flag	A flag to denote if the computed change in course page views is positive	Binary

Table 4.3 continued from previous page

Variable	Description	Type
Appended resources view frequency	Total number of times a student viewed the resources and assignments specifications within the course module from week 1 up to the current week	Quantitative
Resources view frequency	Total number of times a student viewed the resources and assignments specifications within the course module in the current week	Quantitative
Resources views trend	The trend of resources and assignments specifications view pattern over time; data sequences are computed on a weekly basis starting from week 1	Qualitative
Resources views trend flag	A flag to denote if the computed resources and assignments specifications view trend value is positive	Binary
Change in resources views over 1-week	The change in viewing resources frequencies between current and prior week; if the specified week is 1, this variable will be equal to the week 1 resources views frequency	Quantitative
Change in resources views over 1-week flag	A flag to denote if the computed change in resources views is positive	Binary
Total resources opened	Total number of resources and assignments specifications opened by the student within the course discussion forum starting from week 1	Quantitative

Table 4.3 continued from previous page

Variable	**Description**	**Type**
Appended posts view frequency	Total number of times a student viewed the post within the course discussion forum from week 1 up-to the current week	Quantitative
Posts view frequency	Total number of times the student viewed the posts within the course discussion forum in the current week	Quantitative
Posts views trend	Trend of post view patterns over time; data sequences are computed on a weekly basis starting from week 1	Qualitative
Posts views trend flag	A flag to denote if the computed posts view trend value is positive	Binary
Change in posts views over 1-week	The change in viewing post frequencies between the current and prior week; if the specified week is 1, this variable will be equal to week 1 post view frequencies	Quantitative
Change in posts views over 1-week flag	A flag to denote if the computed value of the change in post views is positive	Binary
Total number of read posts	Total number of posts read by student in the course discussion forum starting from week 1	Quantitative

Table 4.3: VLE activities independent variables.

Discussion Forum Variables

Online discussion boards are popular communication and collaboration media. A range of qualitative and quantitative independent variables are extracted from the discussion forum dataset. Qualitative analysis may help to unveil hidden aspects of the discussion forum, such as its social structure and the emotions students express in their posts, alongside other social attributes, reflecting the students' engagement levels. An overview of the independent variables extracted from the discussion forum datasets can be found in Table 4.4.

Variable	**Description**	**Type**
Appended out-degree	(Appended student contribution) Total number of times a student created posts in the course discussion forum between week 1 and the current week	Social
Out-degree	(Current week student contribution) Total number of times a student created posts in the course discussion forum in the current week	Social
Change in out-degree value over 1-week	The change in weekly out-degree value between current and prior week; if the specified week is 1, this variable will be equal to week 1 out-degree value	Quantitative
Change in out-degree value over 1-week flag	A flag to denote if the computed value of change in out-degree is positive	Binary
Appended in-degree	(Appended student posts' popularity) Total number of posts replying to student posts in the course discussion forum between week 1 and the specified week	Social

Table 4.4 continued from previous page

Variable	Description	Type
In-degree	(Current week student posts' popularity) Total number of posts replying to student posts in the course discussion forum in the specified week	Social
Change in in-degree value over 1-week	The change in weekly in-degree value between the current and prior week; if the specified week is 1, this variable will be equal to week 1 in-degree value	Quantitative
Change in in-degree value over 1-week flag	A flag to denote if the computed change in in-degree value is positive	Binary
Appended degree	Appended sum of student engagement in-degree and out-degree in the course discussion forum from week 1 up-to the current week	Social
Degree	Sum of student in-degree and out-degree engagement in the course discussion forum in the current week	Social
Change in degree value over 1-week	The change in weekly sum of in-degree and out-degree values between the current and prior week; if the specified week is 1, this variable will be equal to the week 1 sum of in-degree and put-degree values	Quantitative
Change in degree value over 1-week flag	A flag to denote if the computed change in total-degree value is positive	Binary
Degree trend	Trend analysis of total-degree variables over time; data sequences are computed on a weekly basis starting from week 1	Qualitative

Table 4.4 continued from previous page

Variable	**Description**	**Type**
Degree trend flag	A flag to denote if the computed value of the total-degree trend is positive	Binary
Closeness centrality	The shortest path length from a student to all other participants in the course discussion forum based on posts created between week 1 and the current week	Social
Betweenness centrality	The total number of times in which participants' shortest paths pass through students based on posts created between week 1 and current week	Social
Degree prestige	The ratio of students (popularity) in-degree between week 1 and the specified week	Social
Total discussion created	Total number of discussions created by the student in the course discussion forum from week 1 up-to current week	Quantitative
Posts' sentiment strengths	Sum of student sentiments expressed in the posts and weighted using corresponding adverb strengths in the current week	Qualitative
Posts' sentiment strengths flag	A flag to denote if the computed values of posts' sentiment strengths are positive	Binary
Appended posts' sentiment strengths	Appended sum of posts' sentiment strengths variables based on student posts created between week 1 and the current week	Qualitative

Table 4.4 continued from previous page

Variable	**Description**	**Type**
Appended posts' sentiment strengths flag	A flag to denote if the computed values of the appended posts' sentiment strengths are positive	Binary
Posts' sentiment strengths average	Average of posts' sentiment strengths values in the current week	Qualitative
Posts' sentiment strengths average flag	A flag to denote if computed value of posts' sentiment strengths average variable is positive	Binary
Appended posts' sentiment strengths average	Average of posts' sentiment strengths values based on student posts created between week 1 and current week	Qualitative
Appended posts' sentiment strengths average flag	A flag to denote if the computed value of the aggregated posts' sentiment strengths average variable is positive	Binary
Posts' sentiment strengths trend	Trend analysis of posts' sentiment strengths values; data sequences are computed on a weekly basis based on the student posts created between week 1 and current week	Qualitative
Posts' sentiment strengths trend flag	A flag to denote if the computed value of the posts' sentiment strengths trend variable is positive	Binary

Table 4.4: Discussion forum independent variables.

VLE Access Patterns Variables

Students' usage patterns for VLE in terms of time, expose some aspects of student online learning engagement. Several independent variables have been constructed to explore daily and weekly VLE access patterns. A student is considered connected in an individual day or week when they log-in to VLE at least once in the underlying day or week, based on the VLE log datasets. A description of the VLE access pattern variables can be found in Table 4.5.

Variable	**Description**	**Type**
Count of disconnected days in a week	Count of the number of days in the current week on which the student has not accessed the VLE platform	Quantitative
Count of connected days over 2 weeks	Count of the number of days in a specified fortnight on which student has accessed the VLE platform at least once a day	Quantitative
First connected day in the week	Flags the first day of the current week on which the student has used the VLE platform	Quantitative
Connected flag	A flag to denote if student has logged-in to the VLE platform at least once in the current week	Binary
Total disconnected weeks	Total number of weeks in a row on which the student has not assessed the VLE platform	Quantitative

Table 4.5: VLE access patterns, independent variables.

4.2.4 Dependent Variable

This thesis aims to identify students who are at academic risk by utilising a combination of student online participation data within blended learning courses. Therefore, it is important to specify the circumstances in which a student is considered to be at-risk academically. In this study, we use the final course achievement as a criterion to determine academic risk. A binary dependent variable was computed, where students achieved a final mark over 55 their score is labelled as success academically, whilst a final mark of or below 55 is labelled as at-risk status. In addition, students who choose to drop-out of their courses were coded as having at-risk status, as they did not complete courses successfully. Although the student's final awarded mark of or between 50 and 55 would still pass the course, they would inevitably achieve a final mark that would be relatively close to the failure grade. A total of 1,476 student grades were labelled, where 790 students (53.5 percent) were tagged as being at academic risk and 686 students (46.5 percent) were labelled as having a successful academic status.

4.2.5 Outliers Handling

The VLE participation data were drawn from heterogeneous groups of students. Therefore, it is common to observe outliers, which occur due to variations amongst the students. An outlier is a rare data point compared with other observations that may reduce containing variable productivity power. The outlying cases detection procedure relies on a method proposed by Tukey (1977). Tukey's method identifies outliers through boxplots. The method divides the data into four quartiles, then identifies outliers as observations more than 1.5 times the interquartile range from the quartiles (IQR), which means above (Quarter 3) + (1.5 IQR) or below (Quarter 1) - (1.5 IQR). Outliers were detected based on weekly timeframes and handled prior data analysis. Outliers were detected based on weekly timeframes and handled prior data analysis by removing outlying datapoints from dataset.

4.2.6 Variables Transformation

Data transformation approaches are used to optimise the initial population prior to analysis. In this thesis, min-max normalisation and logarithmic variance-stabilising transformation methods are applied to the students' independent variables. Data normalisation minimises the effects of high observations by outweighing other smaller observations in the variable. The min-max normalisation method performs linear transformations on the initial independent variable data. It rescales processed attribute values in the range from 0 to 1. Min-max normalisation is performed using the following formula:

$$x' = \frac{x - min(x)}{max(x) - min(x)} \tag{4.1}$$

Where:
x refers to the initial value
x' refers to the normalised value
$min(x)$ refers to the minimum value in the x range
$max(x)$ refers to the maximum value in the x range

The other data transformation method is logarithmic variance-stabilising, which is performed using the natural logarithm of the sum of the observation value +1. Extra value is added to each observation, since many observations contain values of 0, especially variables hold actions counts such as the frequency of resources view. For examples, a post view variable will have a value of 0 if a student has never viewed any post in the discussion forum. There is no negative value across the computed variables because of the application of the min-max transformation method.

4.3 Digital Adverb Strength Dictionary

In this thesis, students' sentiments, as they are expressed in the online discussion forums, are used as predictors of their academic risk. Detected sentiments are weighted

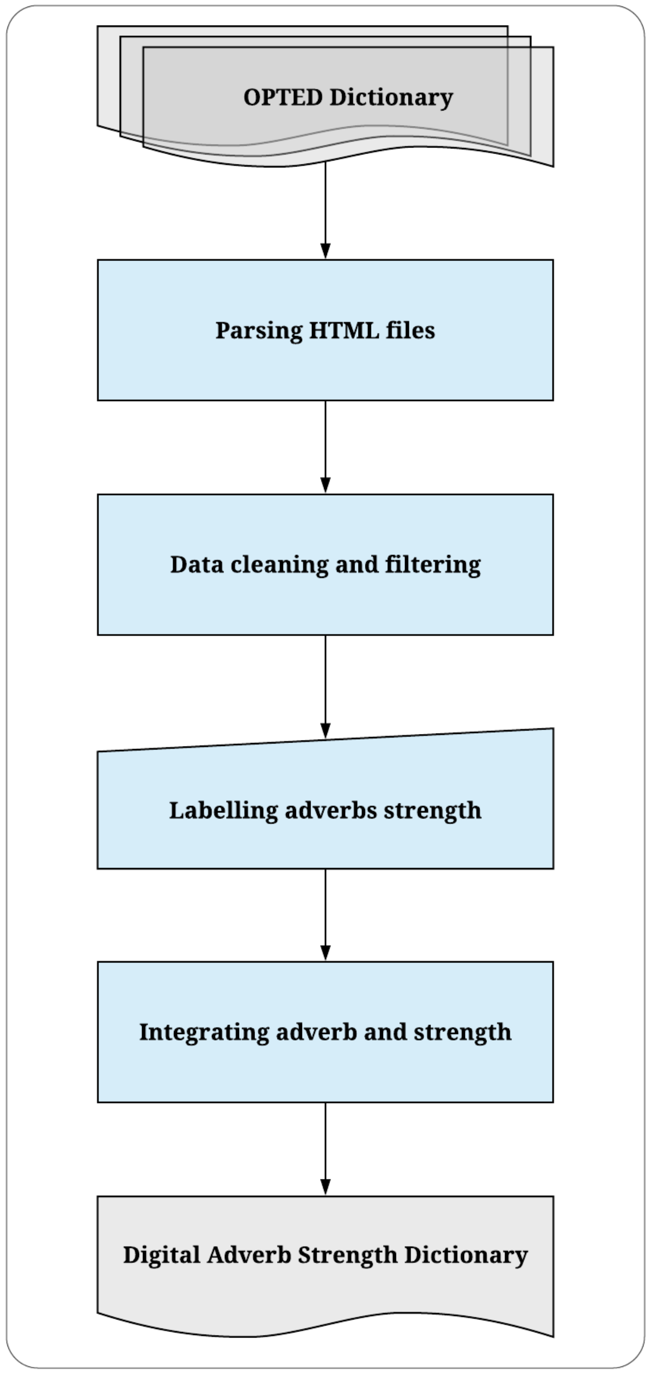


Figure 4.3: Digital Adverb Strength dictionary, building process.

using the strength of the accompanying adverb. The weighting process requires an instrument to identify the strength of each adverb in English. However, to the best of our knowledge, there is no such weighting instrument of adverbial strengths available for use. Therefore, we constructed a Digital Adverb Strength dictionary, where every adverb is associated with its strength. In the dictionary that has been developed, adverbial strength levels are assigned to values ranging from 1 to 3. The process of building the dictionary involved both manual and automated tasks. Figure 4.3 shows the process of building a Digital Adverb Strength dictionary.

The initial dictionary's textual content was drawn from the Online Plain Text English Dictionary (OPTED) version 0.03 (*The Online Plain Text English Dictionary (OPTED)* 2000), which is a public, digital-based, words set. OPTED provides a set of 26 HTML files that contain lists of English words, accompanied by the words' definitions and part of speech. The dictionary's datasets were parsed and cleaned by removing HTML tags, word definitions and parts-of-speech labels. Furthermore, all non-adverb arguments were filtered out and omitted from the dictionary dataset. Then, the adverbs' strengths were labelled manually, where the strongest adverbs have a strength level of 3 and weaker ones have a strength level of 1. The constructed adverb strength dictionary was stored in an XML file to be used in this study. It contains 3,762 weighted adverbs.

4.4 Stanford Natural Language Processing Toolkit

Natural Language Processing (NLP) is a collection of methods that aim to analyse and understand human-generated texts. Natural Language refers to human language in use. NLP applies mathematical and artificial intelligence approaches to textual-based elements, to determine the author's mood or the emotion expressed in a piece of writing. There are a variety of NLP applications in many sectors, such as mining customers' opinions about a product or tracking people's attitudes on social media.

Sentiment analysis is one commonly used application of NLP, where it enables determination of individual sentiments, expressed in a piece of text. Numerous NLP systems provide a variety of automated annotators, such as sentiment analysis and part of speech tagging. The Stanford CoreNLP Natural Language Processing Toolkit is a well-known example of such computer-based NLP tools.

CoreNLP is an automated annotation-based NLP tool developed by the Stanford Natural Language Processing Group at Stanford University. It is one of most popular NLP tools (Manning et al. 2014). The CoreNLP toolkit has had three releases, which were implemented in the Java language. Although the first version of CoreNLP was developed for internal use, to replace an older system, in 2010, version 3 was released as a free open source package, since it was licensed under the GNU General Public License on version 3 or later (*Stanford CoreNLP Natural language software* 2015). It supports multiple languages including Arabic, Chinese, English, French and German. CoreNLP offers various essential NLP annotations such as sentiment analysis, part of speech tagging and syntactic analysis.

This study utilises NLP to detect sentiments expressed in discussion forum posts and identify weighted sentiments using associated adverbs. For this purpose, we employed the Stanford CoreNLP toolkit for NLP tasks. The Stanford CoreNLP version 3.2.1, English set, was used in this thesis. CoreNLP is responsible for identifying sentiments in students' posts containing textual content and tagging the associated adverbs, which are then weighted using the adverb dictionary mentioned earlier. The CoreNLP sentiment annotator reports expressed sentiments as having integer values on five levels between 0 and 4, where 0 indicates a strong negative sentiment and 4 indicates a strong positive sentiment. However, we rescaled the initial produced values to be between -2 and 2, as the following negative values refer to negative sentiments; 0 refers to natural, positive values where the mean positive sentiment was detected, and the opposite is true.

4.5 Machine Learning

Machine learning involves a range of algorithms and approaches that allow computer systems to make successful predictions by observing the relationships between input variables and prediction target. A more formal definition of machine learning is provides by Mitchell (1997) as "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E".

Machine learning filed provides approaches that can robotically learn from input observations to build a computational predictive model. The process of building predictive models is known as "training" while input data used to build predictive model is called "training data".

In supervised machine learning methods, learning techniques requires the target value for each training instance to be provided. Learning approach trains underlying predictive models to map the relationship between the target values and input observations.

The rest of this section provides an overview of machine learning approaches that was utilised to construct the predictive models used to determine those students at academic risk. Furthermore, it discusses Ensemble-based modelling approach which has been used throughout the thesis.

4.5.1 Logistic Regression

Regression analysis is defined as a process involving several techniques for forecasting the relationship between the response variable and single or multiple explanatory variables. While regression analysis has its roots in the field of statistics, it has commonly been used in the field of machine learning. Regression models are powerful prediction approaches. Consequently, it is one of the most regularly used prediction techniques in many scientific fields (Armstrong 2012). Nevertheless, this prediction technique produces more reliable estimations when it deals with small numbers of variables, and big amounts of data, where changes are larger and more predictable and there are strong causal relationships (Armstrong 2012). In machine learning, a variety of regression techniques are used to predict continuous or categorical targets, while others forecast binary outcomes such as logistic regressions. A logistic regression is a generalised linear model that is usually applied to fulfil binary classifications in its simple form. On the other hand, multinomial logistic regression models can solve multiclass classification tasks.

In educational research settings, a wide range of previous efforts utilise regression models to predict course outcomes. For example, Ashenafi et al. (2015, 2016) employ linear regression models to predict students' final examination scores. Logistic regression is a popular model for binary predictions targets. Many studies rely on logistic regression-based models to predict students who are at-risk of academic failure (Jayaprakash et al. 2014, Macfadyen & Dawson 2010, Marbouti et al. 2016, Mueen et al. 2016, Pardo et al. 2016), while others utilise the same approach to predicts students who are at-risk of attrition (Burgos et al. 2018, Chai & Gibson 2015, He et al. 2015).

Since we are targeting predicting students' academic status as either successful or at-risk, we employ a logistic regression technique, as its predictive power is illustrated in the literature. Several previous works in the area of predicting student academic risk, show that logistic regression models can produce reliable predictions and they outperform many other approaches, when compared within a prediction setting. Logistic regression algorithms report on probability values, based on the given independent variables, whenever the probability of an above-threshold student is considered as have a successful academic status, otherwise they are considered as having an at-risk status. The logistic regression function is expressed mathematically using the following equation:

$$p = \frac{1}{1+e^{-(\alpha+\beta_0 X_0+\beta_1 X_1+\ldots+\beta_p X_p)}} \tag{4.2}$$

4.5.2 Ensemble Modelling

The concept of combining multiple predictive models' outputs is known as ensemble modelling in machine learning context. Machine learning ensemble-based models are constructed by combining multiple classication approaches to enhance the predictive accuracy over a single learning model (Dietterich 2000). An ensemble model consists of a number of learning members which generalise the capability of predictive instrument to combine various hypotheses from the hypothesis space over single weak learners.

Bagging and boosting are well-known ensemble methodologies in machine learning (Kantardzic 2011). Fundamentally, bagging and boosting approaches are based on resampling training examples. For instance, bagging method build learning members by training each learner on a random sampling of training instances where each learning member is generated with different random sampling of training instances. In terms of boosting ensemble method, it tends to used entire training dataset with updating training instance weights after every development iteration.

Finally, applying a suitable outputs combination strategy is a vital element of building ensemble models, as it assigns the contribution degree of each learning member on the final model output. Weighting models' outcomes is an example of the blending strategy used to combine models' outputs in ensemble modelling (Polikar 2012). Assigning weights to members allows us to manage members' degrees of importance and contributes to the final decisions. For instance, higher weightings are given to strong member classifiers, which results in higher contributions to the model's outcome.

Non-trainable and trainable weighting methods are used to allocate weight parameters associated with each single model prediction (Polikar 2006). Non-trainable combination rules allow the user to specify the weight parameters to be applied to the member models. On the other hand, a trainable weighting approach recognises the models' weights through a training algorithm, where it optimises a best-fit set of weights that produces the best performance.

4.6 Evaluation Methods

Assessing the quality of predictive instruments is a critical task in the process of developing new models. A variety of evaluation metrics have been used to evaluate the performance of different machine learning approaches, which allow the researcher to examine the predictive model's effectiveness and compare it with other models in terms of predictivity power. This study targets binary classification, where the final output is that either the student is at-risk or has a successful academic status. The rest of this section presents an overview of popular evaluation mechanisms used to evaluate the performance of binary classifiers.

A typical approach used to evaluate the performance of machine learning models is based on a confusion matrix. Figure 4.4 presents a confusion matrix for a binary classifier. Several commonly-used evaluation metrics are calculated based on a confusion matrix for binary classifiers. Traditionally, student success models rely on the overall accuracy, precision, recall and F-measure metrics to examine the model's overall quality and accuracy. However, the overall accuracy may lead to misleading evaluation results in the case of an evaluation dataset containing highly imbalanced samples. Recently, many studies have utilised the area under the ROC curve metric to assess the performance of binary classification tasks, especially in the case of imbalanced datasets. This section presents frequently-used evaluation methods.

Where:
True Positive (TP) refers to the number of students classified as at-risk correctly

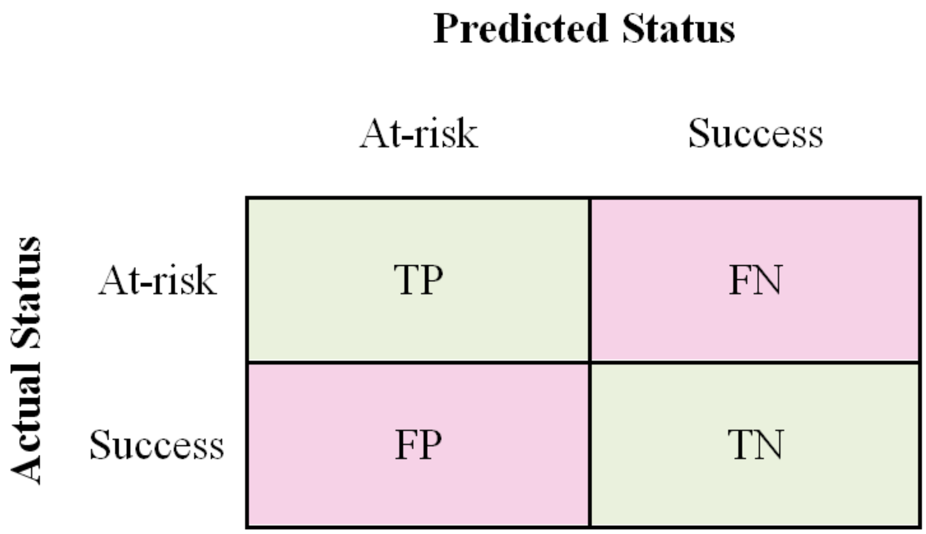


Figure 4.4: Confusion matrix for a binary classifier.

True Negative (TN) refers to the number of students classified as successful correctly
False Negative (FN) refers to the number of actual at-risk students misclassified mistakenly
False Positive (FP) refers to the number of actual successful students misclassified mistakenly

Accuracy measures the overall correctness of the classifier, as follows:

$$Accuracy = \frac{True\ Negatives + True\ Positives}{Negatives + Positives} \tag{4.3}$$

Precision (that is, the Positive Predictive Value) focuses on the proportion of samples labelled positive, classified positive correctly, as follows:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{4.4}$$

Recall (that is, Sensitivity) focuses on the proportion of actual positive samples, classified positive correctly, as follows: Recall= (True Positives)/(True Positives+False Negatives)

$$Recall\ = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{4.5}$$

Specificity focuses on the proportion of actual negative samples, classified negative correctly, as follows:

$$Specificity\ = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \tag{4.6}$$

$\mathbf{F_1}$ **score** measures the harmonic mean between precision and recall:

$$F_1\ score = 2\ \frac{precision \cdot recall}{precision + recall} \tag{4.7}$$

Area Under the Curve

The ROC (Receiver Operating Characteristic) curve results from the plotting sensitivity against the 1-specificity at numerous threshold points. The AUC (Area Under the Curve) computes the area under the ROC curve. The AUC is a beneficial method for evaluating and comparing binary classifiers.

4.7 Summary

This chapter provides a detailed description of the collected data used in this thesis, which consists of records of students' online engagements with course VLE and discussion forums. Furthermore, it presents the methods used for data preparation prior to analysis, including data cleaning, time-series clustering, outlier handling and data transformation methods. Moreover, the set of 53 student level predictors used in this study are listed and described alongside descriptions of all dependant variables.

The chapter also provides an overview of the tools and approaches used to extract and weigh the sentiments students express in their posts. Furthermore, the development process for the Digital Adverb Strength dictionary and its uses are presented in this chapter. In addition, contextual notes are given for the logistic regression technique and ensemble modelling approach that is used to develop predictive models for student success. Finally, the chapter presents the evaluation metrics that are used to ensure the quality and accuracy of the prediction instrument developed to identify at-risk students.

The next chapters present the use of the prepared student data, with the help of the approaches and tools described above, which are then used to develop and evaluate a multi-course, adaptive early warning framework of students at academic risk. While Chapter 5 describes and evaluates the proposed discussion forum predictors alongside the proposed novel Grey Zone design for improving binary classifiers, Chapter 6 presents the development and evaluation of an exemplar multi-course early warning framework, which integrates discussion forum predictors alongside other VLE interaction variables for early identification of at-risk students in Computer Science courses. In Chapter 7, we extend our work by developing an adaptive mechanism, which allows the multi-course early warning framework to enhance its performance and ability to detect at-risk students across different educational settings, alongside coping with any changes that may occur in the prediction space over time. Adaptive mechanisms allow the underlying predictive instrument to learn robotically from any extra dataset which may become available in the future.

Chapter 5

Early Detection of At-Risk Students Using Course Discussion Forum Data

5.1 Overview

VLEs supports a variety of online educational tools including online communication spaces such as online discussion forums. Online forums are asynchronous web-based platforms, which allow students and lecturers to communicate virtually, without time or physical limitations (Loncar et al. 2014). Online discussion forums have gradually become an important part of computer-supported courses. Virtual discussion forums optimise the learning process by providing a space for students to seek and receive help, as well as become involved in course-related discussions with lecturers and other peers outside lecture times and campus borders.

On the other hand, online discussion forums enhance the lecturers' ability to monitor students' learning progress by analysing discussion forum activities, digital traces and contributions to textual content. Several studies found positive relationships between students' participation in virtual communication environments and academic performance in a higher education setting. For example, in (Cheng et al. 2011), (Gunnarsson

& Alterman 2012) and (Shaw 2012), researchers observed positive links between several students' engagement characteristics in learning communication platforms and academic achievements. A variety of discussion forum factors were utilised to predict students' academic performance. However, the majority of the previous work mainly relies on quantitative measures of participation frequency and social network aspects to predict students' academic achievements.

While multiple studies observed an association between content-based aspects and student performance in a higher education context (i.e. (Tucker et al. 2014, Wen et al. 2014*a*)), this type of feature is rarely used to predict academic performance. Language-based qualitative analysis involves utilising a range of NLP approaches, such as identifying students' sentiments expressed in their discussion forum messages. In this chapter, we focus on evaluating the ability of discussion forum predictors to predict students who are at academic risk. Hence, we extract various aspects of students' engagements with online forums by performing quantitative, qualitative and social measures to explore hidden aspects of students' participations. Furthermore, in this study, we propose an automated approach, weighting student sentiments based on the strength of associated adverbs by multiplying the sentiment value computed from the CoreNLP toolkit with the strength of the accompanying adverb gained from the Digital Adverb Strength Dictionary. The feasibility of the proposed language-based features, alongside other predictors, is evaluated in terms their importance to the predictive instrument for an at-risk student.

Then, we employed extracted predictors to develop an early predictive instrument that was fed with discussion forum data only to predict students' academic status (at-risk or successful). Given the binary nature of the classification targets, various binary classifiers have been used to fulfil prediction tasks in higher education literature. However, traditionally, researchers follow black-and-white decision-making strategies, where, if the computed probability falls below a certain threshold, the student is predicted to be at-risk, otherwise the student is allocated to the class of successful students.

In this chapter, we propose a novel, to the best of our knowledge, design of Grey Zone decision-making and prediction strategy for binary classifiers, where the computed probabilities are divided into three zones (black, grey and white). In the proposed design, instances which fall within the Grey Zone bounds are subject to further investigation, using additional predictive models, to distinguish their actual class.

To examine the effect of the Grey Zone design on the performance of a binary predictive instrument, we built twelve different models, trained with temporal information, where each model corresponds to an individual lecture week. We ran a series of experiments to evaluate the performance of predictive models over traditional and proposed designs. Then, we evaluated the impact of the proposed design on the overall predictive model performance across each prediction week, individually.

The rest of this chapter is organised as follows: Section 5.2 describes the data used in this chapter. Section 5.3 presents discussion forum features and their extraction methods alongside the proposed automated sentiment weighting approach used to evaluate the strength of students' posts. It also describes the feature ranking approaches used to prioritise features based on their importance level. Furthermore, Section 5.4 describes the proposed novel Grey Zone strategy for binary classification. It also describes the experimental setup used to evaluate the impact of the proposed design over commonly used designs, alongside presenting the predictive models' development process and specifications. Sections 5.5 and 5.6 show and discuss the experimental results, respectively. Finally, Section 7 summarises the chapter.

5.2 The Dataset

Underlying data were collected from thirteen Computer Science courses taught at the University of Adelaide, Australia, between 2012 and 2016, drawn from the Moodle VLE as described in Section 4.1. The original dataset covers 1,476 students, enrolled in thirteen courses. The dataset contains a total of 3,797 posts created by students, lecturers and tutors, since students' contributions in course-based online discussion forums are the central element used to predict students' overall course performance in this chapter. Therefore, only students who posted at least once during a relevant semester's official academic calendar were considered in this chapter. However, non-student participants contributions are kept in the dataset to maintain information about students' social ties with other users. Overall, the cleaned dataset contains data belong to 451 enrolments across the thirteen collected courses that represents approximately 30 percent of the gathered population.

5.3 Discussion forum Features Experimental Setup

5.3.1 Weighting Sentiment Approach

Textual content posted on online discussion forums is a rich source of information. Such data may contain valuable information reflecting hidden characteristics of students' learning experiences, including sentiments regarding course-related aspects that can indicate their academic progress. Therefore, it is central to utilise such a source of information to improve the ability to predict students' academic performance.

Sentiment analysis refers to the process of classifying emotions and attitudes expressed in texts as positive, natural or negative (Liu 2011). Sentiment analysis applies NLP approaches to investigate lexical elements of a piece of text that is examined using machine learning or lexical-based approaches. In predicting tertiary student academic performance setting, sentiment analysis is a rarely used method to explore student-generated texts, related to course contexts. The extracted information can be used as a predictor of student academic risk behaviour. A few researchers have utilised sentiment analysis as a predictor of student retention or final achievements in a higher education setting such as in (Wen et al. 2014*a*,*b*).

In this chapter, we propose an automated weighing mechanism to rank students' sentiments, as they are expressed in the online discussion forums. In the developed mechanism, detected sentiment polarities are weighted using the strength of the accompanying adverbs, where the strength of each adverb is identified using a Digital Adverb Strength Dictionary built for this purpose. To examine the practicability of the proposed feature extraction approach in predicting students' performances, the predictive power of the weighted sentiment features were compared against other discussion forum characteristics.

Digital Adverb Strength Dictionary

To perform a sentiment weighting approach automatically, a digital dictionary of adverbial strengths is required to define the strengths of English adverbs. However, to the best of the author's knowledge, there is no such weighting of adverbs strength dictionary currently available for use. Therefore, a Digital Adverb Strength dictionary has been constructed, where every adverb is associated with its strength. The digital dictionary categorises English adverbs into three strength levels ranging from 1 to 3. The process of building the dictionary involved manual and automated tasks, as described in Section 4.3.

5.3.2 Automated Sentiment Extraction and the Weighting Process

In this thesis, NLP tasks are processed with the help of the Stanford CoreNLP toolkit version 3.2.1, English set (*Stanford CoreNLP Natural language software* 2015). CoreNLP is responsible for identifying sentiments expressed in post textual contents and tagging associated adverbs, which are weighted using the Digital Adverb Strength dictionary. The CoreNLP sentiment annotator reports the expressed sentiment polarity indicator as an integer value on five levels between 0 and 4 where 0 indicates a strong negative sentiment and 4 indicates a strong positive sentiment. However, we rescaled the initially produced indicators to be between -2 and 2 for clarity, as the negative values refer to negative sentiments; 0 refers to natural; positive values mean a positive sentiment was detected and the opposite is also true. Figure 5.1 shows a workflow diagram of identifying and weighting sentiments in online discussion forum.

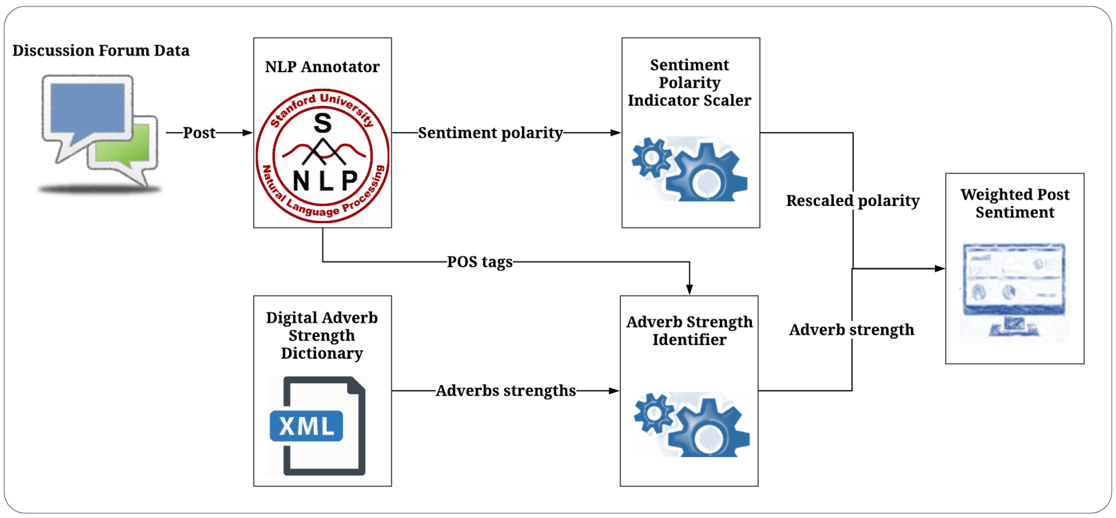


Figure 5.1: A workflow diagram showing posts' sentiment extraction and weighting.

5.3.3 Features Description

In order to provide early predictions of academic risk, student features were calculated based on their contributions to course discussion forums on a weekly basis. Several qualitative, quantitative and social analytic methods were used to compute the emotional, participation and social aspects, alongside measuring the change in weekly observations over the semester's duration for each student. Therefore, two data clustering criteria were used to compute the variables which are: single week clustering criteria and weeks appended criteria. Single week features are computed based on activities performed in the prediction week only, while appended features combine actions performed from study week 1 up to the underling week. Table 5.1 presents the extracted independent discussion forum features used in this work.

Feature	Analytic Method
Appended posts' sentiment strengths average	Qualitative
Appended posts' sentiment strengths average flag	
Appended posts' sentiment strengths	
Appended posts' sentiment strengths flag	
Degree trend	
Degree trend flag	
Posts' sentiment strengths	
Posts' sentiment strengths flag	
Posts' sentiment strengths average	
Posts' sentiment strengths average flag	
Posts' sentiment strengths trend	
Posts' sentiment strengths trend flag	
Appended degree	Social Network Analysis
Appended in-degree	
Appended out-degree	
Betweenness centrality	Social Network Analysis
Centrality degree	
Closeness centrality	
Degree prestige	
Degree	
In-degree	
Out-degree	
Change in degree value over 1-week	Quantitative
Change in degree value over 1-week flag	
Change in in-degree value over 1-week	
Change in in-degree value over 1-week flag	
Change in out-degree value over 1-week	

Table 5.1 continued from previous page

Feature	Analytic Method
Change in out-degree value over 1-week flag	
Total discussion created	

Table 5.1: A list of the extracted discussion forum features.

Qualitative analysis approaches focus on evaluating sentiments expressed in posted messages' contents weighted by their strength by performing the weighting sentiment method described earlier. Posts' textual contents are evaluated using the post sentiment weighted by the associated adverbs' strengths automatically with the help of *Stanford CoreNLP Natural language software* (2015) and the Digital Adverb Strength dictionary developed in this study. In addition, other qualitative methods were used to evaluate the fluctuation degree of underlying student sentiments over time based on weekly observation blocks.

Furthermore, quantitative measures apply statistical methods to calculate student participation characteristics. Extracted indicators involve a frequency analysis of students' participation in discussion forums, including the frequency of creating discussion threads by each student. Moreover, quantitative features measure the change in the number of posts created and replies received, compared with the previous study week.

Additionally, social network analysis examines social ties within discussion forums and relationships between members. A number of social network analysis indicators were extracted including centrality, closeness and betweenness degrees, and students' prestige degrees. Social network analysis examines the popularity, importance and centrality of a student within the social network.

5.3.4 Feature Ranking Methods

In a machine learning setting, utilising feature selection methods is a popular practice to select the most relevant and influential subset of features (Hira & Gillies 2015). Features selection algorithms evaluate the influence of each individual feature in the features space and rank them, based on their importance to the prediction model.

Multiple feature selection approaches were used to prioritise the importance of the extracted discussion forum factors in predicting students' risk behaviours. The process of ranking features' importance involves feeding all underlying data to multiple feature ranking approaches including: the information gain approach, which measures the amount of information that that a particular feature has about each class, the gain ratio, which applies similar measurements, such as information gain, but whilst reducing its bias by considering the number and size of the data, Pearson's correlation, which examines the linear correlation between each feature and its class and the ReliefF technique, which scores features by distinguishing qualities between classes. Features ranking is performed with the help of a Weka machine learning library (Frank et al. 2016).

All the available student data was cleaned and pre-processed, then aggregated to be fed to each feature selection approach. The pre-processing task covers handling outlying data points and transforming the computed observations, which aims to optimise the initial population prior to analysis. Two data transformation methods were performed on independent features, which are the min-max normalisation and logarithmic transformation. Each data pre-processing task was performed each week, for each course dataset, separately.

5.4 Grey Zone Design Experimental Setup

5.4.1 Grey Zone Design

A wide range of machine learning approaches have been used to predict future events. However, algorithms are varied with respect to their prediction outcomes. Several prediction models target predicting continuous and categorical classes, while others target binary outputs.

In a binary classification context, researchers customarily follow black-and-white decision-making strategies. To illustrate this, in a case where the probability computed by the prediction instrument falls below a threshold, the instance under prediction is assigned to a certain target class, otherwise the instance is allocated to other classes in the prediction space, which forces the prediction instrument to make marginal decisions regarding the prediction class. Hence, traditional decision-making methods may limit the model's ability to provide quality predictions when underlying features reflect similar observations for instances belonging to different target classes. Therefore, we proposed the novel design of a Grey Zone strategy to improve the accuracy of binary classifiers, by re-predicting instances in the Grey Zone using an alternative model. The Grey Zone strategy focuses on identifying the weakness of the predictive model where instance characteristics overlapped, and the model produced random predictions by analysing misclassification composition to identify the Grey Zone. In other words, the Grey Zone is intended to cover a range of probabilities where the predictive model fails to provide accurate predictions.

In the proposed design, a base model is used to calculate the preliminary probabilities. Then, the base model output distribution is analysed and divided into three zones: the black, white and Grey Zones. The Grey Zone refers to a range of probabilities that need further investigation, while the black and white zones are assigned directly to one of the target groups. The Grey Zone's upper and lower boundaries are stated with the help of an error analysis of the base model classifications, using a receiver operating characteristic (ROC) curve graph and visualisation of the computed probabilities distribution. Error analysis is performed to identify the potential Grey Zone, where a high percentage of misclassifications occurs.

Appling error analysis of the base model allows to identify best-fit global threshold. Then, we identify optimal cut-off values for the upper and lower boundaries of the Grey Zone with respect to the base model global threshold. At this stage, we estimate the upper and lower cut-off values (the optimal Grey Zone upper and lower thresholds) based on a Receiver Operating Characteristics (ROC) graph (Fawcett 2006). ROC works by drawing many points on the graph space starting from the lower left point (0,0) to the top right point (1,1). For starting and finishing points, the predictive model predicts instances to a single class unconditionally, while the upper left point (0, 1) characterises the finest classifications. Therefore, measuring the distances between each point in the ROC space and the top left point indicates the best cut-off point where the shortest distance is the best cut-off point. Two cut-off values are identified with respect to global threshold (upper and lower Grey Zone boundaries) by detecting optimal cut-off values with respect to the area above and below the global threshold.

To overcome the predictive model weakness, the Grey Zone model may consider different aspects of the instance characteristics, expand the feature coverage, utilise different training strategies and/or follow alternative modelling approaches to optimise the final classification quality. Then, any instances which fall within the Grey Zone boundaries are subject to additional investigation using the Grey Zone predictive model, which replaces the initial prediction output produced by the base model. Figure 5.2 presents an overview of the Grey Zone design architecture.

The concept of combining multiple predictive models' outputs is known as ensemble modelling in machine learning context. The fundamental concept of ensemble modelling id used to develop the proposed Grey Zone design. However, the design utilises a novel development approach and design comparing to well-known ensemble

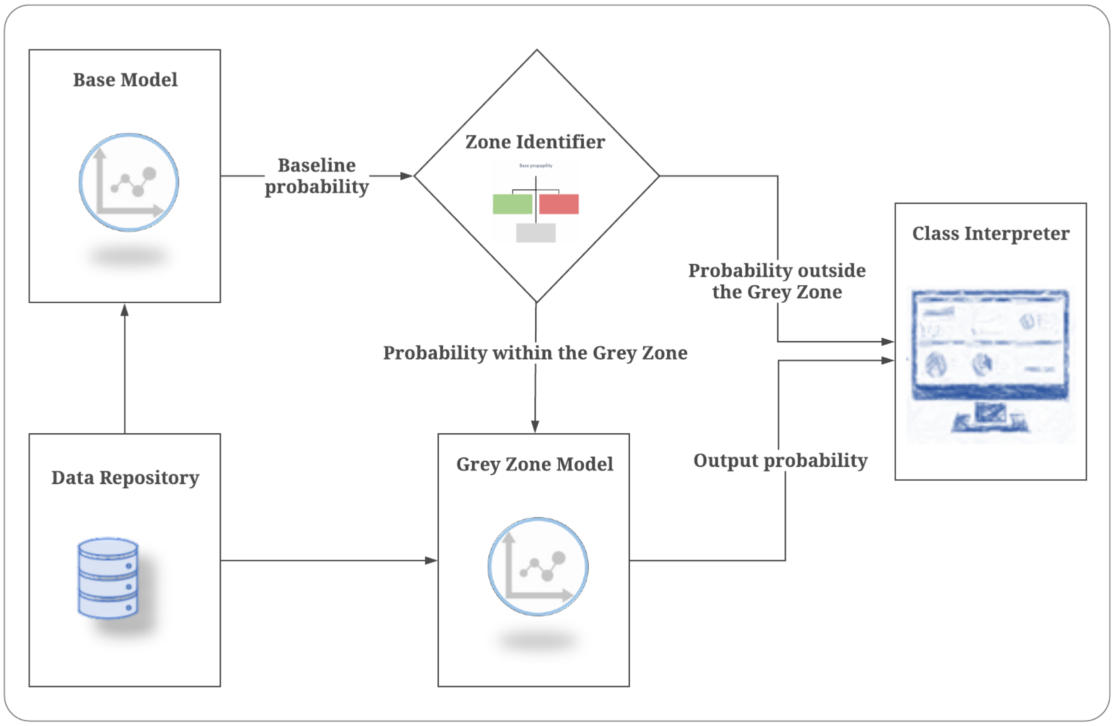


Figure 5.2: Grey Zone design architecture.

modelling approaches such as boosting approach. Although the proposed Grey Zone design and boosting approach tends to optimise the overall performance of predictive ensemble model when the base learning member provides unstable predictions, but each approach built based on a different methodological concept.

Initially, the Grey Zone design is strictly limited to only two members, baseline and grey zone models, while boosting approach can involve N number of members. Moreover, boosting approach incrementally building new members based on the latest developed member performance, however, the grey zone design involves the process of identifying the Grey zone based on the performance of a static baseline model. Furthermore, in terms of utilising training dataset, boosting approach tends to used entire training dataset with updating training instance weights after every development iteration, on the other hand, in the grey zone design only instances fall in the identified grey zone used to develop grey zone member model. Finally, in boosting approach, the outcomes of all members combined to frame final outputs while in the proposed design the grey zone models' outcomes replace baseline models' outcomes whenever under prediction instances allocated within the identified grey zone, otherwise baseline outcome used.

In this chapter, we evaluate the impact of the concept of the Grey Zone on the performance of the predictive instrument in the context of predicting students who are at-risk. The purpose of utilising the proposed concept is to distinguish students' actual classes when students behave similarly by employing additional features in the Grey Zone model, thereby optimising the quality of the final predictive model by enhancing its ability to provide more accurate predictions of both at-risk and successful student classes.

5.4.2 Prediction Strategy

This work aims to provide weekly predictions of the students' academic risk status throughout the 12-week long semester, alongside evaluating the impact of the Grey Zone design on the overall prediction quality and accuracy over traditional black-and-white strategy. For each study week, a set of online discussion forum participation predictors are extracted using students' involvement data, performed on current prediction week only and aggregating the data up to the current prediction week, where the prediction target is expressed as a binary prediction task (at-risk or successful). The extracted predictors are prepared and pre-processed to be fed into the predictive models. Binary logistic regression is used to estimate the probability of a student being at-risk academically. Experiments were conducted to predict students' academic performances at two stages. The first stage involves developing weekly predictive models using traditionally-used decision-making strategies that are also used as base models in the Grey Zone design. The second experimental stage applies the Grey Zone concept, including identifying the best-fit Grey Zone boundaries and developing Grey Zone models. The model-building process was carried out using version 3.8.2 of the Weka machine learning library (Frank et al. 2016).

- **First stage:** Baseline models are built and evaluated following a commonly-used, black-and-white decision-making strategy. Twelve models were developed, where each model corresponds to each study week, and implemented to predict those students who are potentially at academic risk. Each predictive model was fed with a dataset containing temporal-based training information extracted from the online discussion forum contributions preformed on or before each week's end. The base models' performances were validated using an independent testing dataset.

- **Second stage:** The Grey Zone design was utilised to examine its impact on the weekly models' quality. The models built in the first stage presented base models in the proposed design, where the prediction outcomes are analysed weekly to determine the upper and lower boundaries of the Grey Zone. Then, the Grey Zone models were built to differentiate between at-risk and successful students when they behave similarly. The exact same training and testing datasets utilised in the first stage were used to feed and evaluate the Grey Zone models.

Before building each model, we performed a features selection task to select the most influential subset of predictors from the features space. The wrappers feature selection approach (Kohavi & John 1997) is utilised to fulfil the feature selection task. The wrappers approach is carried out by training the predictive models with different combinations and subsets of features, followed by a comparison of the models' performances. The model evaluation process involves utilising AUC and overall accuracy metrics to measure and compare the models' qualities and performances.

Another important aspect that must be considered is the proportion of students who are labelled as either at-risk academically or successful in the training dataset. In the utilised dataset, the percentage of instances that are labelled as at-risk is about 25.5

percent, which considers such students as a minority. Any class imbalance in the training dataset leads to negative effects on the classifier's performance. There have been many attempts to address this problem, usually by over-sampling the minority class or under-sampling the majority class, such as (Gray et al. 2016, Jishan et al. 2015, Thammasiri et al. 2014), in the context of predicting student academic performance. Over-sampling and under-sampling schemas increases and decreases respectively the occurrence of instances belong to one target class in the training dataset to enhance the performance of the predictive subject model. Sampling schemas are popular in the machine learning context, as these methods only modify the imbalanced training dataset and do not require changes to be made at the prediction algorithm level.

However, in the underlying set of experiments, we handle the data imbalance problem by adjusting the threshold for each model to the best performing value, rather than by using the most-used threshold value of 0.5 as needed. In the case of the utilised unbalanced training dataset, thresholding is used to avoid a severely high misclassification rate in one class, where lowering the threshold leads to an increasing specificity rate, while raising threshold value results in a higher sensitivity rate regarding the subject's predictive model. In other words, the threshold adjustment method allows for training models with imbalanced datasets, while remaining capable of producing accurate predictions over minority classes with overall good accuracy. To compute the adjustment threshold value for each weekly model, we utilised an ROC graph, which can provide insightful assessment regarding the model's performance at each possible threshold point in the probability range.

Building Base Models

In a binary classification setting, commonly such models are utilised to predict target classes where instances are assigned directly to one of outcome classes. However, in the proposed design, the base models generate baseline probabilities where instances fall in the Grey Zone probabilistic range and are then subject to further investigation. Otherwise, instances that have computed a likelihood outside the Grey Zone are allocated directly to the relevant prediction class. Therefore, it is important to build accurate models, as the Grey Zone concept relies primarily on their outcomes to distinguish students' classes. Throughout the development process, the performances of weekly base models are evaluated following the black-and-white concept where best-performing models were nominated.

Each week's base model is fed with a corresponding temporal dataset that contains data logged from the beginning of the semester up to the underlying model's week. Although temporal modelling approaches reduce the size of the training dataset, it involves greater advantages, such as avoiding the effects of high observation points that arise towards the end of the courses. A range of features extracted from online discussion forum participants' data were computed using quantitative, qualitative and social approaches to build these models.

The wrapper feature selection approach was performed to reduce the feature space by selecting the most powerful subset of predictors to build the models, taking into the account different sizes of training datasets available for training each week's model. The wrapper approach selects the most relevant set of features for the prediction concept and avoids noise features from the full features space. The features selection step for the base models involves selecting a fixed set of features to be used across the weekly base predictive models. Instances belonging to the 12 prediction weeks were aggregated into a single dataset to perform the features selection task. The selected subset of features made up of a combination of 10 features was used to build base models across all the prediction weeks where each weekly model is fed with the corresponding temporal training dataset. Table 5.2 lists a subset of features selected to develop the base model.

Feature	**Category**
Appended degree	
Appended in-degree	
Degree	Participation

Table 5.2 continued from previous page

Feature	**Category**
In-degree	
Out-Degree	
Total discussion created	
Centrality degree	Social
Appended posts' sentiment strengths	
Posts' sentiment strengths	Posts' weighted sentiments
Posts' sentiment strengths trend	

Table 5.2: A list of the selected subset of features used to build the main predictive model.

A related point to consider is that the training dataset contains an imbalanced class distribution where the minority instances belong to the at-risk prediction class. In order to overcome this problem, we adjusted the threshold for each model with the help of an ROC graph to control the threshold adjustment value for each weekly model.

Identifying Grey Zone Boundaries

An initial step to utilise the proposed Grey Zone strategy is identifying the optimal cut-off values that separate the subject zones within the prediction model outcomes range. Identifying the Grey Zone's upper and lower boundaries was performed by analysing the prediction errors arising in the weekly base model estimations. The first step was to visualise the distribution of the calculated probabilities produced by the base model for each week. This step aims to provide a preliminary idea about the area where a high portion of probabilities overlapped across all the weeks. The overlapped area shows where misclassification occurs for both the target classes. Figure 5.3 presents a visualisation of the outcome probabilities distribution for testing week 8.

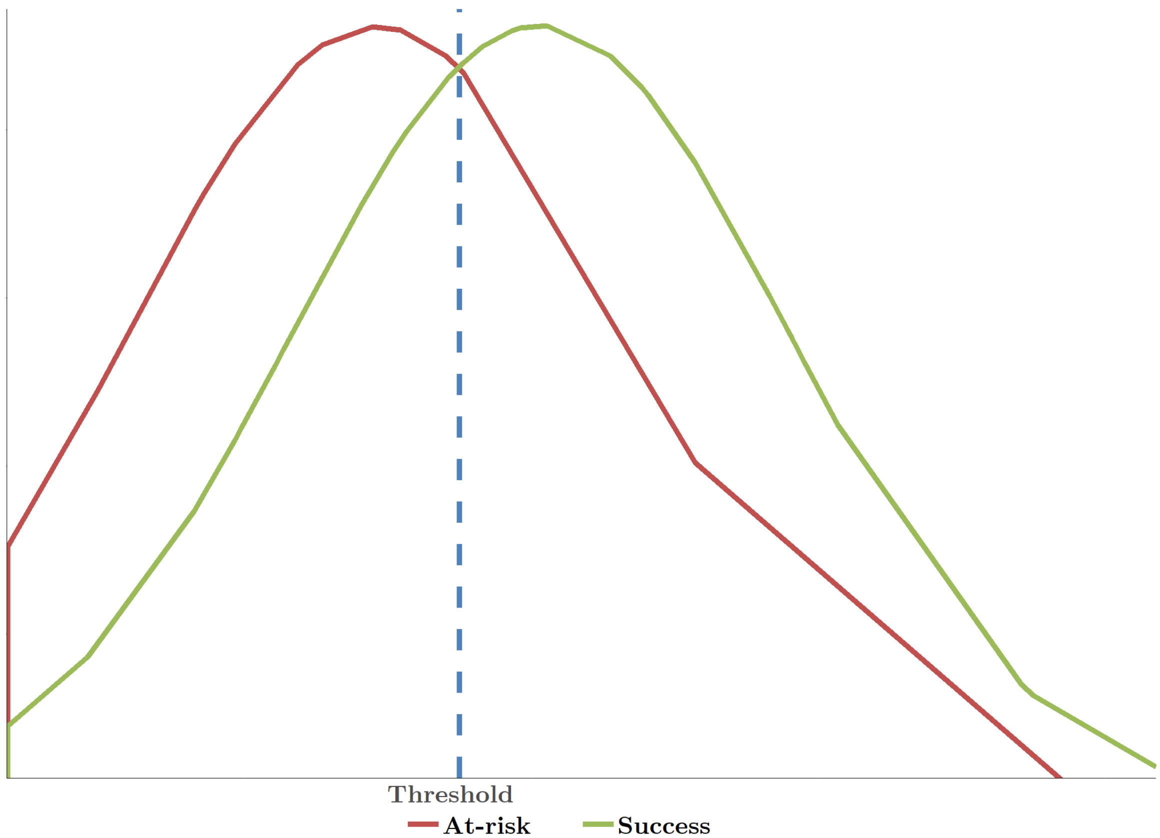


Figure 5.3: A visualisation of the probabilities distribution produced by the week 8 base model.

The second step was performed with the help of an ROC graph. The ROC graph was utilised to reveal the models' performances at each possible threshold point in the prediction space. An ROC graph helps to identify the best-fit cut-off points, which reepresent the optimal Grey Zone boundaries. A pair of upper and lower Grey Zone boundaries was generated to accompany each weekly base model.

After performing an error analysis, we nominated the value of the threshold at + 0.15, that being the value upper boundary and the value of the threshold at - 0.15 as being the lower boundary of the Grey Zone for all weekly base models. The identified Grey

Zone has a high misclassification rate, where over 92 percent of overall misclassifications occur within its boundaries across all weeks. The developed base models are able to provide quality predictions above and below the Grey Zone boundaries. The Grey Zone design assigns instances which are allocated in white or black zones to be the successful or at-risk prediction classes respectively. However, instances that have base model estimations that fall within these boundaries are subject to re-prediction using Grey Zone models.

Grey Zone Models

The purpose of Grey Zone models is to complement the base models. Grey Zone models are designed to overcome the weakness in the base models, where they fail to distinguish actual instance classes. Since instances which have similar participation characteristics regarding the utilised range of features in the base models have consequently similar estimation values, it is important to design the Grey Zone models carefully to handle this situation and to enhance the final outcomes' quality and accuracy. Only instances which have a base model probability falling between Grey Zone boundaries are subject to further investigation by the Grey Zone models.

To develop the best performing Grey Zone models, two different Grey Zone ensemble models were developed, where each ensemble model corresponds to a set of study weeks. A Grey Zone ensemble model is designed specifically for study weeks 5 to 8, as they surround the mid-semester break, where students' engagement characteristics change significantly. The other model covers the rest of the semester.

Ensemble modelling involves combining a set of models' predictions, which allows us to combine various hypotheses from the hypothesis space. Each ensemble model consists of two logistic regression members alongside the underlying week base model developed previously, while the majority voting approach is used to combine the three models' predictions, where instances allocated to the class have more votes. Although Grey Zone member models mainly rely on features driven from the current week's participation and features reflecting the change and fluctuation in participation patterns, each one of the member models is considered an alternative design to achieve the best prediction performance.

The first member model is built using the same training approach used to build the base models but extends the range of utilised features to form the features space. The second member model is fed with the semester-aggregated dataset to increase data coverage by training models with the total available dataset. Wrapper feature selection approaches were utilised to select the most powerful features associated with Grey Zone instances for both newly developed member models. Different sets of features were selected for each of the ensemble model. Table 5.3 lists the subsets of features used to build the two ensemble models.

Feature	**Category**
Appended degree	
Appended in-degree	
Change in degree value over 1-week	
Change in in-degree value over 1-week*	
Change in in-degree value over 1-week flag*	
Change in out-degree value over 1-week*	
Change in out-degree value over 1-week flag*	Participation
Degree	
Degree trend	
Degree trend flag	
In-degree	
Out-Degree	
Total discussion created	
Betweenness centrality	Social
Closeness centrality	
Appended posts' sentiment strengths	Weighted Sentiment

Table 5.3 continued from previous page

Feature	Category
Appended posts' sentiment strengths flag**	
Posts' sentiment strengths	
Posts' sentiment strengths average*	
Posts' sentiment strengths trend	

Table 5.3: Subsets of features selected to build Grey Zone ensemble models. * Feature is used only in Grey Zone models corresponding to weeks 1-4 and 9-12. ** Feature is used only in Grey Zone models corresponding to weeks 5-8.

Moreover, due to the imbalanced class distribution of the utilised dataset, we adjusted the threshold for each weekly Grey Zone member model to reduce the error rate in the models' members. The threshold adjustment process is utilised by plotting an ROC graph for each Grey Zone member model with its corresponding testing instances to determine the best performing threshold value.

5.5 Results

5.5.1 Features Importance Ranking

Given our ultimate goal is to develop a reliable and timely predictive framework of academically at-risk students, in this chapter we explored numerous predictors drawn from online forums data, including social and participation characteristics alongside predictors that were computed based on the sentiment weighting approach and trending analysis of students' weekly participation and contribution patterns. It is crucial to evaluate the importance of each predictor to identify which online forums factors have a heavy influence on the prediction outcomes.

A preliminary analysis is conducted to compare two sets of language-based predictors to determine the influence of proposed sentiment weighting mechanism on predictors over non-weighted sentiment predictors. The first predictor set computed using only sentiment values identified by the CoreNLP toolkit while the second predictor set built by weighing extracted sentiment with the help of the proposed sentiment weighting mechanism. To compere the sets of predictors, we ranked language-based predictors using ve well-known feature selection approaches that are Pearson's correlation, information gain, the gain ratio and ReliefF feature selection approaches using all available datasets. Ranking analysis shows that most of language-based predictors build using sentiment weighting mechanism ranked higher corresponding predictors built using sentiment identified by CoreNLP toolkit including appended posts' sentiment strengths, appended posts' sentiment strengths average, posts' sentiment strengths and posts' sentiment strengths average. Overall, the proposed sentiment weighting approach resulted in improving the quality of language-based predictors. Therefore, the proposed mechanism is used to construct language-based predictors in this study.

To identify the most influential predictors, we ranked the predictors as described in Table 5.1, based on their importance level to the predictive model, reflecting their significance to the final predictions. Discussion forum predictors' importances were ranked using five well-known feature selection approaches. The ranking methods include Pearson's correlation, information gain, the gain ratio and ReliefF feature selection approaches. The extracted weekly features were aggregated in a single dataset and fed to each one of the utilised ranking approaches to evaluate each predictor's importance throughout the semester.

Overall, features driven from the linguistic analysis of discussion forum messages present the majority of the top ranked predictors across all the proposed approaches. Table 5.4 illustrates the highest five features resulting from each ranking approach. The semester-aggregated sum of the weighted posts sentiments is the best predictor among the underlying discussion forum predictors. It was selected as having the

Feature	Pearson's correlation	Information Gain	Gain Ratio	ReliefF
Appended posts' sentiment strengths	3	**1**	**1**	**1**
Posts' sentiment strengths trend	-	-	-	3
Posts' sentiment strengths trend flag	**1**	2	3	-
Posts' sentiment strengths	-	-	-	4
Posts' sentiment strengths average	-	-	-	5
Betweenness centrality	5	5	2	-
Degree trend	4	3	4	2
Degree trend flag	2	4	5	-

Table 5.4: Top 5 features ranking list where the highest ranking features have a rank of 1.

top-ranking position by three ranking methods and was the third ranking by Pearson's correlation approach. The posts' sentiment strengths trend flag is the second overall best ranking, where it indicates the direction of students' sentiments over the semester. The third best predictor is computed based on trend analysis of student collaboration patterns, followed by its flag indicator. Only one social characteristic appears in the top 5 ranking list, which is the betweenness centrality. Finally, three additional language-based features ranked among the best 5 features at least once, all stemming from the ReliefF approach. In sum, the feature ranking results present the valuable influence of students' sentiment factors on predicting at-risk students in a higher education context.

5.5.2 Prediction Performance, Applying the Grey Zone Design

In this chapter, we developed series of models to predict student performance based on discussion forum data in a weekly manner. A set of experiments have been carried out to examine the robustness of such a data source to provide early predictions of student academic risk behaviours. In the initial experimental stage, we predicted each student's academic status using base models only, while in the subsequent experiments we applied the Grey Zone approach on the same set of weekly models to measure the impact of the Grey Zone's design on the overall models' predictivity. To evaluate the performance of the developed models, we performed a course-based splitting approach, where populations belonging to the same course must be fully assigned to either training or testing datasets. While dividing the dataset, we paid attention to maintain a similar distribution of the sample classes (at-risk and successful) across the training and testing datasets. The training dataset contains the largest portion of the population size since it is the foundation of the proposed models. A total of 395 students were assigned to the training dataset which contained those students who had enrolled in 11 courses. The testing dataset contained 56 students belonging to two courses. Although the weekly models utilise the exact same set of features, the amount of data made available for the training models was gradually increased every week as a result of the following temporal training approach.

Before evaluating the models' performances, it is important to select a suitable evaluation metric. Given the binary nature of the problem and the problem of the unbalanced distribution of classes in the gathered dataset, we utilised AUC metrics to examine the quality of each model, alongside employing an overall accuracy based on the confusion table for each model, as it provides an overview of the overall classification accuracy.

In the first experimental stage, we utilised a traditionally-used decision-making approach where the model outcome distribution was split into two regions belonging to either at-risk or successful classes. We built a predictive model to predict students' who are at-risk. The weekly predictive performance ranges between 0.51 and 0.73 in terms of the AUC metric, while the models achieve accuracy ranging from 35.7 percent to 71.4 percent. However, the model performance fluctuates throughout the semester as models' performances vary each week. The model achieved its peak AUC value in week 8 and the highest accuracy at week 10.

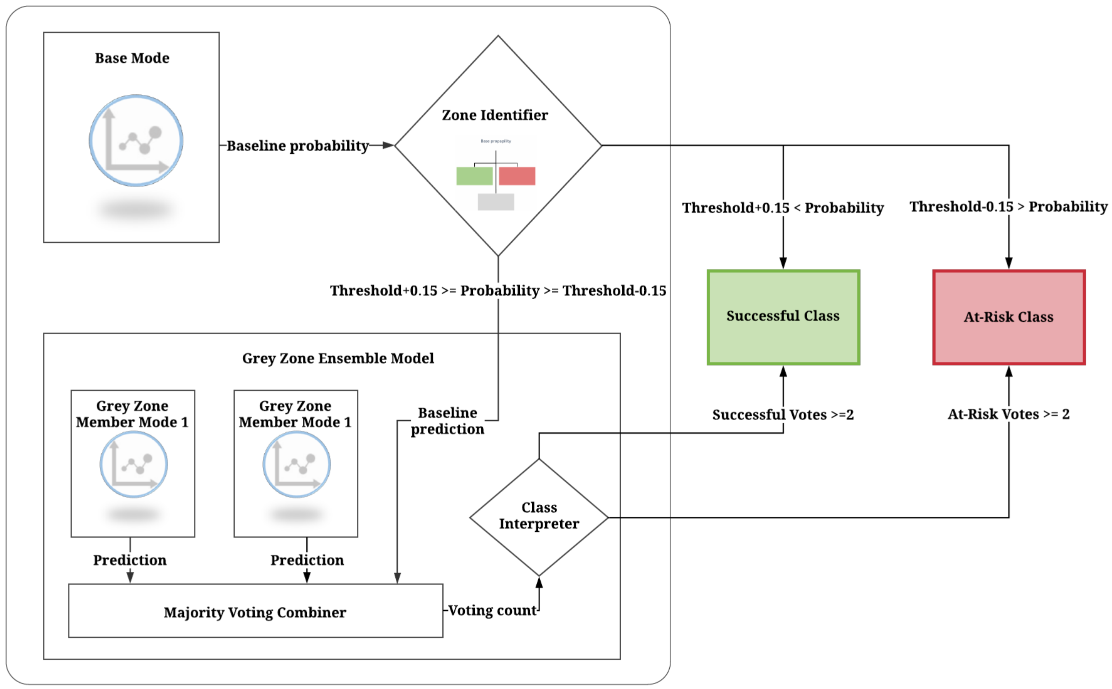
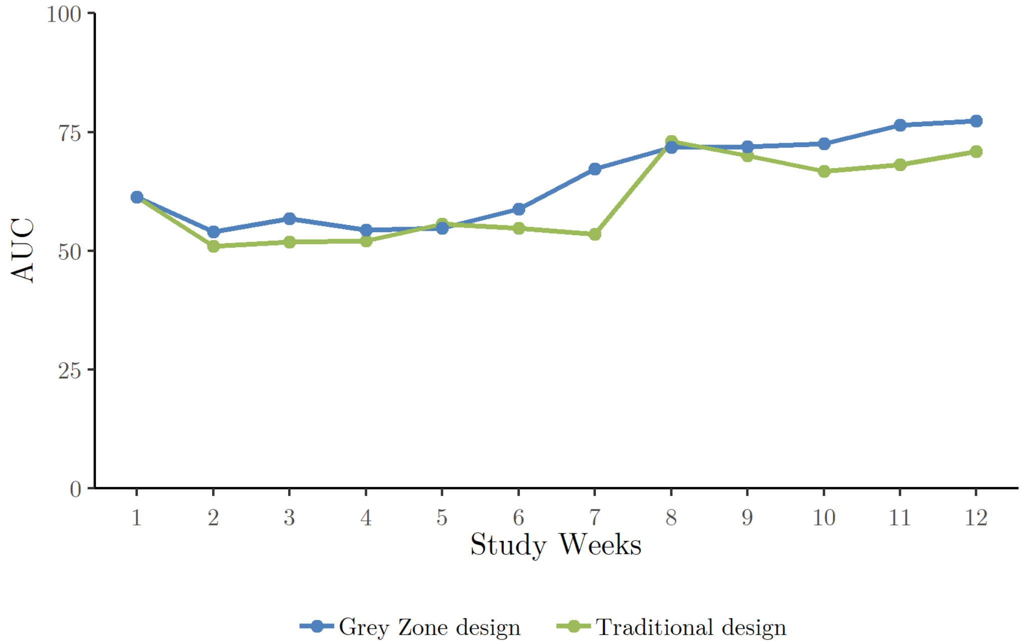


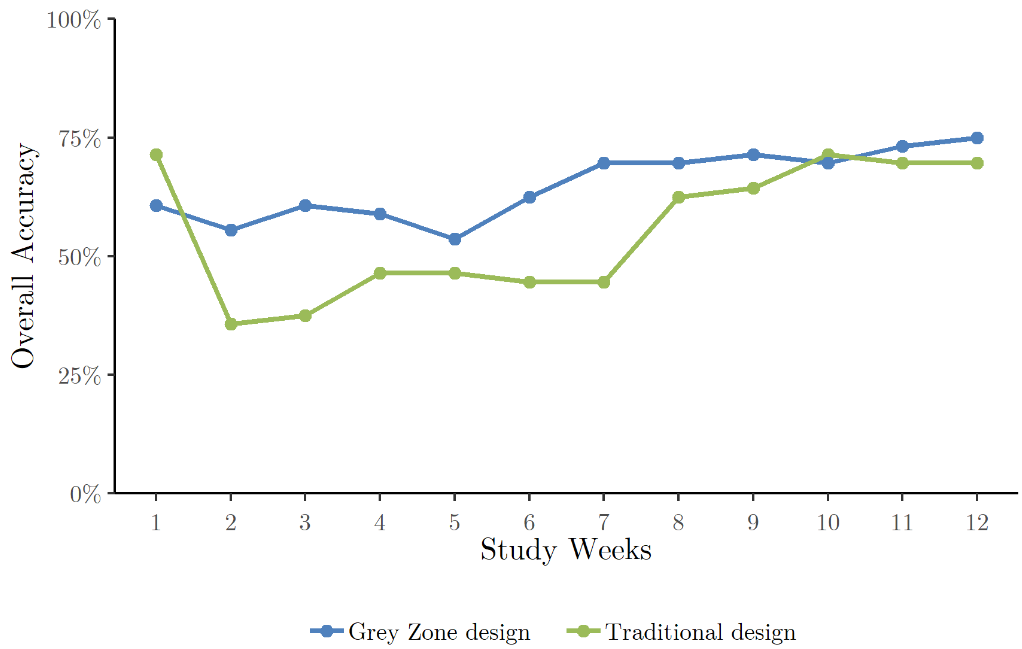
Figure 5.4: The structure of predictive models using the Grey Zone design.

Then, we extended our work by applying the Grey Zone design to combine the Grey Zone models to each corresponding weekly model. Figure 5.4 shows the structure of the utilised weekly predictive model following the Grey Zone design where the predictive models developed in the first experimental stage are used as base models at this stage. In the proposed design, the base model computed probabilities that fall above or below the identified Grey Zone boundaries are predicted as successful or at-risk, respectively. Two Grey Zone ensemble models were developed, where one ensemble model is associated with study weeks near the mid-semester break and the other model is concerned with the remaining study weeks. Utilising Grey Zone design results in achieving weekly AUC values ranging between 0.54 and 0.77, while achieving an overall accuracy ranging from 53.6 percent to 75 percent.

Table 5.5 and Figure 5.5 demonstrate the weekly models' performances in terms of AUC and the overall accuracy across the two modelling approaches. Generally, integrating Grey Zone models improved the performance of the base model throughout prediction weeks. Utilising Grey Zone design helped to improve the quality of the weekly models predictivity throughout the semester, while the predictive models achieved their best performance at the end of the semester. When comparing the impact of Grey Zone models over base-models only on each individual weekly model, the proposed Grey Zone design enhances the models' quality across most prediction weeks. However, the performance of the prediction models following traditional modelling in study weeks 5 and 8 outperformed the performance of the proposed Grey Zone modelling slightly in respect of AUC due to their relative closeness to the mid-semester break, where week 5 is one week prior to the break and week 8 is one week after the break. Although the Grey Zone design reduces overall accuracy in weeks 1 and 10, it resulted in enhanced prediction quality in week 10 and improved the accuracy of the rest of the weeks by about 13 percent on average.



(a) AUC



(b) Overall accuracy

Figure 5.5: Weekly models' performances over the traditional and Grey Zone modelling designs.

	Traditional design		Proposed design	
	AUC	**Accuracy**	**AUC**	**Accuracy**
Week 1	0.61	71.4%	0.61	60.7%
Week 2	0.51	35.7%	0.54	55.5%
Week 3	0.52	37.5%	0.57	60.7%
Week 4	0.52	46.4%	0.54	58.9%
Week 5	0.56	46.4%	0.55	53.6%
Week 6	0.55	44.6%	0.59	62.5%
Week 7	0.54	44.6%	0.67	69.6%
Week 8	0.73	62.5%	0.72	69.6%
Week 9	0.70	64.3%	0.72	71.4%
Week 10	0.67	71.4%	0.73	69.6%
Week 11	0.68	69.6%	0.76	73.2%
Week 12	0.71	69.6%	0.77	75.0%

Table 5.5: A demonstration of the weekly models' results over the commonly used and proposed Grey Zone modelling designs where, in the proposed design, base models are exactly the same models used in traditional modelling experiments.

In sum, with the help of Grey Zone modelling, the models' performance reached up to 0.77 in terms of the AUC metric and a peak overall accuracy of 75 percent with a higher recall and precision average. Utilising the Grey Zone concept allows predictive models to produce noticeably overall better quality predictions in the early study weeks. The Grey Zone design improves the weekly models' performances by up to 5 percent in terms of the AUC metric and increases prediction accuracy by up to 23 percent in the first quarter of the semester. Furthermore, the design has a significant positive impact on the majority of the models' performances throughout the remaining study weeks, where it was able to enhance the weekly models' predictive power by up to 13 percent and up to 25 percent in terms of AUC and overall accuracy measures, respectively.

5.6 Discussion

Online discussion forums provide virtual space for students to facilitate their learning experience, as they promote collaborative learning and communication between students and both their peers and their lecturers. Students' contributions in online discussion forums can be used as an indicator of their final course performance. However, in the context of predicting students' performance in higher education, most of the previous work has focused on utilising quantitative analysis of students' actions performed on social online platforms. In this chapter, we evaluated the predictivity of discussion-forum-related predictors to forecast university students who are at-risk academically on a weekly basis, in blended learning setting.

Discussion forums are online platforms that are used for course-related discussions for which participation are typically voluntary. Optional involvement in course forums may lead to low engagement rates in many cases. However, other factors may also impact on engagement levels, such as the course structure (Yukselturk 2010) and number of enrolments in the course (Vrasidas & McIsaac 1999). Hence, it would be useful to explore additional aspects of the available data rather than relying solely on counting the number of created and received posts. A relatively few studies utilised social aspects (i.e. (Haig et al. 2013, Jiang et al. 2014, Romero et al. 2013)) and linguistic aspects (i.e. (Tucker et al. 2014, Wen et al. 2014*a*)) of course discussion boards to predict potential at-risk students.

In this chapter, we explored various characteristics of discussion forum participation data by performing content-based, social and frequency analyses alongside analysing the changes in contributions produced by each student in a weekly manner throughout the semester. In order to determine the most influential factors' effects on students who might be at-risk, we began by evaluating and ranking the predictivity power of individual discussion forum factors. Predictors of evaluation results show that the best ranked predictor is related to students' sentiments expressed in their posts that are weighted using the proposed adverbs strength method, followed by the social betweenness centrality degree and features reflecting the fluctuation in participation patterns over the study weeks.

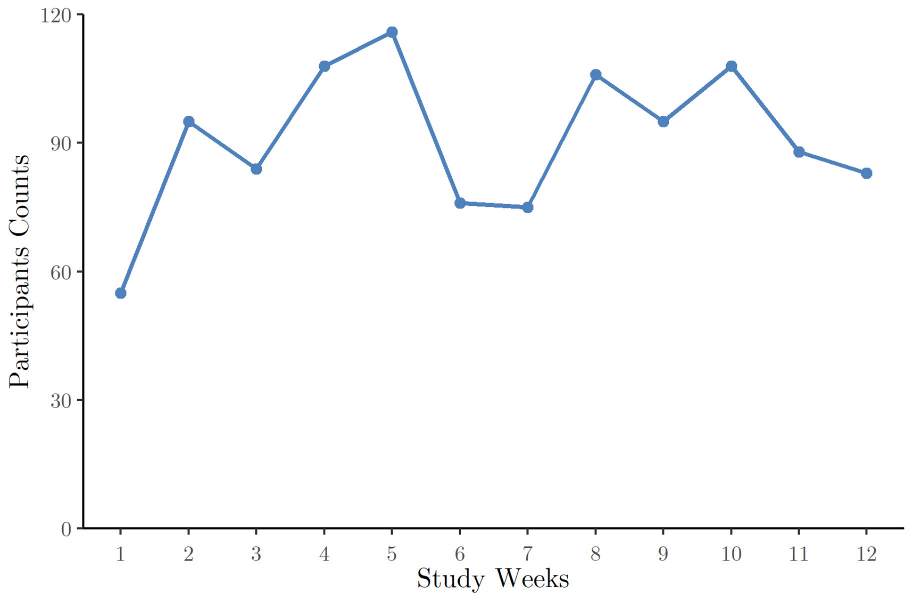
On the other hand, in the process of developing a predictive instrument, depending on a single or a small bag of predictors may cause type I and type II errors to arise, as these predictors cover only one or limited aspects of students' academic risk factors. Moreover, the predictors have different contribution levels on the models' outcome, therefore, in this chapter, we built weekly computational predictive models using combinations of the most influential features to achieve the best possible quality and accuracy.

Early identification of at-risk students allows lecturers to support students in a meaningful and timely manner. It is important to provide accurate results for both prediction classes (at-risk and successful) without assigning higher importance to one class. This derives from the fact that misclassifying at-risk students mistakenly as successful may lead to a delay in supporting those students who are in need, while in the opposite case, lecturers are subject to extra workload as the number of students labelled as at-risk increases.

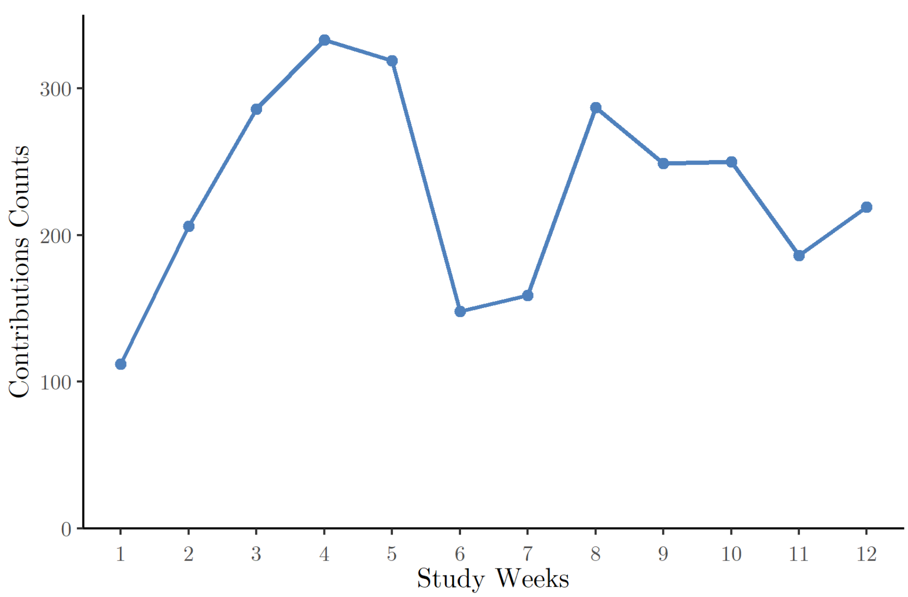
We performed a binary classification task, as existing studies apply a black-and-white strategy to assign a computed likelihood to one class or the other based on a prespecified threshold. However, we then introduced a Grey Zone prediction strategy to improve the overall prediction quality. The proposed design allows for further investigation for students located within the Grey Zone boundaries, where most of misclassifications occur in base model predictions, based on each model's threshold. The Grey Zone models employ different sets of models to distinguish students' classes. The design involves creating a base model and a Grey Zone model corresponding to each week.

The base models were developed using temporal modelling approaches instead of using common semester-aggregated datasets to avoid the effects of high participation records that arise towards the end of the courses. Moreover, to overcome issues related to the imbalanced distribution of classes, we adjusted each model's thresholds to improve the models' predictivity. Error analysis of base models' predictions were used to identify Grey Zone boundaries where the substantial majority of misclassifications arise. Grey Zone models employ ensemble modelling to tackle problems associated with similar participation characteristics by combining a bag of models for each study week. Each Grey Zone model was built using different subsets of features from those used in the base model. Grey Zone models corresponding to most weeks employ the same modelling features. However, special subsets of features were assigned to Grey Zone models corresponding with study weeks five to eight, as the mid-semester break falls in between them and consequently student participation patterns change significantly. Figure 5.6 present the students' weekly engagement in discussion forums in terms of the number of weekly contributions and participants, making visible the changes in student engagement patterns.

Despite the novelty of the proposed Grey Zone design, the experimental results show that applying the Grey Zone strategy resulted in overall noticeable improvements regarding AUC and overall accuracy metrics. The proposed design allows us to correctly identify instances that were misclassified by the base model. Furthermore, it improves the overall stability of the model predictions over the duration of the course. However, the Grey Zone models corresponding to weeks 5 and 8 reduce the quality of the base model slightly in terms of the AUC metric, giving lower accuracy due the high increase in the participation rates of some students, which caused high variability in the participation patterns in the specified weeks. The proposed models started to provide reliable results after the mid-semester break (study week 7), while the models' performance improved towards the end of the semester due the significant increase in the available data at the end of the semester regarding the testing population. The final week model achieved the best performance by classifying 75 percent of the students correctly with a recall value of 81.3 percent.



(a) Number of participants



(b) Contributions rate

Figure 5.6: Students' discussion forum data.

This chapter relies on investigating online discussion forums data alone to predict at-risk students, which highlights the limitation of utilising this data source. Firstly, the use of course discussion forums is usually optional, which makes it impossible to predict the performance of students who choose not to participate in online discussions, instead those students are considered as at-risk even if they have good academic standing. In addition, although the discussion forum data is a rich source of data concerning students' social and emotional standing, it does not reflect the full picture of each student's performance. For instance, in a blended learning environment, the course online forum represents one component among a range of other online and off-line learning activities and communication channels such as communication with lecturers and other students via email or in person. These communication activities are not recorded in the utilised data, which leaves us unable to track them.

Finally, despite the limitations of relying solely on discussion forum participation data to predict student performance, the test results provide evidence about how it can be valuable to explore students' post contents to assess students' risk status. Moreover, the results show that utilising linguistic and social elements alongside other participation factors can provide reliable predictions of students who are at-risk with the help of the Grey Zone design. This design is a promising method to improve the classifiers' quality, as it helped the weekly models to improve their final classification accuracy as well as enhancing the models' ability to distinguish correct instance classes.

5.7 Summary

This chapter focused on mining discussion forum contributions data with the objective of providing early predictions of students who are at-risk of not completing their academic courses successfully. Investigating student-generated textual content, social characteristics and participation patterns within online discussion forums open doors towards unveiling hidden aspects of students' learning experiences. Data was collected from thirteen blended computer science courses offered at postgraduate and undergraduate levels. In the first stage, we evaluated and ranked the importance of students' contributions characteristics on their performance predictive model. The overall ranking results reveal that the appended sentiment strength feature is the most significant discussion forum predictor, followed by their degree of social prestige. The extracted features were used to implement predictive models and provide predictions on a weekly basis.

Furthermore, we evaluated the effectiveness of the proposed novel Grey Zone decision-making design to improve the quality of the binary classifiers. The proposed design suggests further investigation for students for whom their calculated probability falls within pre-defined boundaries. Initial comparison of the experimental results shows that applying the Grey Zone design over the traditional decision-making strategy improves the overall weekly model accuracy by up to 25 percent. Experimenting with the proposed Grey Zone design resulted in providing enhanced overall model quality by providing higher recall and precision on average throughout the 12 lecture weeks.

Finally, in this chapter, we were restricted only to students who have posted at least once in the course online discussion forum to build and evaluate the early predictive instrument. In the next chapter (Chapter 6), we extend our work by utilising additional data courses to generalise the predictive models and improve their quality. The next chapter presents the effects of integrating VLE interactions data with discussion forum data to build an exemplar multi-course framework for early identification of at-risk students. Combining VLE interactions data allows us to explore further aspects of students' interactions with VLE and consequently enhance our ability to distinguish academic risk characteristics early in the semester. Since the proposed Grey Zone approach shows its ability to enhance the model's predictive power, the same prediction concept is taken into consideration when building the predictive model in the next chapter to enhance the framework's performance.

Chapter 6

An Exemplar Multi-Course Early Warning Framework to Identify At-Risk Students in Blended Learning Computer Science Courses

6.1 Overview

Integrating VLEs into the learning process has become a vital practice in modern higher education. Utilising VLEs as an educational tool alongside traditional faced-to-face teaching methods forms the blended learning mode (Garrison & Kanuka 2004). Despite VLE being the main focus of improving the learning experience, it brings challenges related to monitoring and predicting students' learning performance using students' digital footprints. Various studies highlight the relationship between VLE digital traces and final course outcomes in a formal higher education setting. For example, past studies conducted by Agudo-Peregrina et al. (2014) and Cerezo et al. (2016) observed a positive correlation between multiple VLE aspects and course academic performance.

On the other hand, with the dramatic increase in the number of tertiary students and the huge volume of online interactions data, lecturers have become interested in automated tools to analyse online learning behaviours and accordingly predict potentially academically at-risk students. Early identification of struggling students allows lecturers to provide interventions for those students in need in timely manner, which leads to improvements in students' learning experiences and university outcomes (Arnold & Pistilli 2012, Burgos et al. 2018, Cassells 2018, Dodge et al. 2015, Jayaprakash et al. 2014, Smith et al. 2012). Sclater et al. (2016) report several international case studies conducted in Australia, the UK and the US that demonstrate the positive impact of early interventions on students' academic achievements.

Therefore, developing a quality early warning framework to predict potentially at-risk students is a critical step in delivering proactive support to students in need. A variety of students' characteristics gained from different sources were fed into a wide range of approaches to build early warning frameworks. Dynamic data presents popular sources of information that can reflect students' online learning progress. However, relying on data retrieved from VLEs alone to develop an early warning framework is a challenging task.

In higher educational contexts, courses are varied with respect to course structures, required workloads and assessments. Therefore, it is important to consider the diversity in courses, whilst developing predictive instruments for at-risk students. However, most of the existing early warning systems assume that courses have homogeneous data distributions. Hence, in most cases, researchers build and validate early warning systems by splitting a single dataset that is drawn from a single or limited number of courses, without taking into the consideration testing the systems using unseen datasets to anticipate their performance in future courses.

The goal of this work is to build an exemplar multi-course early warning framework that provides quality predictions of at-risk students early in the semester across computer science courses that use blended learning theory as their central pedagogy. The proposed multi-course early warning framework predicts students' academic performance based solely on online learning data extracted from VLE.

Given the binary nature of prediction outcomes, logistic regression is one of the most popular approaches for binary targets in the educational literature (i.e. (Bainbridge et al. 2015, Dominguez et al. 2016, Li et al. 2017)). Traditionally, researchers follow black-and-white decision-making strategies, where, should the base model output fall below a threshold, the student is predicted to be at-risk, and the opposite is true. In this chapter, we utilised the Grey Sone decision-making strategy for binary classifiers proposed in Section 5.4.1, as the strategy illustrates its ability to optimise the overall models' performance in predicting at-risk students. Furthermore, the multi-course predictive framework is built using an ensemble modelling strategy, where eight different members were combined to construct the predictive model. Members were combined in a way that enables a rise in the influence of the quality model members on the final model outcomes, where each member is an expert in a local domain of the features space.

Moreover, the performance of the multi-course early warning framework is assessed using a fresh evaluation dataset drawn from a variety of computer science courses. The evaluation dataset contains unseen course data, where each course has its unique structure and activities distribution. The evaluation process helps to address the methodological gap by examining the framework's ability to predict future events.

An additional aim of this chapter is identifying the optimal week in which course coordinators and lecturers should start intervening to support students in need. The proposed predictive framework aims to provide reliable predictions of at-risk students as early as possible. However, at the beginning of the semester only a limited amount of student interactions and contributions data is available, which reduces the framework's ability to distinguish the students' correct academic status. On the other hand, towards the end of the semester, the framework maintains significant amount of students' information, but late intervention may result in limited intervention impact. Hence, we take both facts in consideration while we evaluate the framework's weekly performance to determine the best time to start offering additional support to students.

The rest of this chapter is organised as follows: Section 6.2 provides a brief description of the multi-course early warning framework for at-risk students. Section 6.3 describes the ensemble modelling design used to develop the predictive instrument used to predict students' academic status. Furthermore, Section 6.4 describes the collected dataset and online learning features used in this chapter. Section 6.5 shows the experimental setup which demonstrates the experiential workflow and framework development process. Finally, Sections 6.6 and 6.7 present the experimental results and provide a brief summary of this chapter.

6.2 A Multi-Course Early Warning Framework for At-Risk Students

The availability of students' demographic and learning characteristics data opens new doors in exploring multiple aspects of academic risk factors in a higher education setting. Early initiatives for utilising digital learning footprints aimed to identify students who are at academic risk in order of support them to enhance universities' outcomes (Campbell et al. 2007). Recent research integrates learning analytics methods in an effort to develop sophisticated predictive instruments that detect at-risk students in terms of retention and/or the risk of underperforming. However, rarely have initiatives have been turned from a raw concept to a developed, operational, predictive instruments. In fact, the majority of the existing work is limited to identifying the most influential predictors of academic risk and comparing the performances of various prediction techniques to distinguish the best performing approach.

In this work, we focused on leveraging virtual learning data to build a multi-course early warning framework for blended learning context. The predictive framework is intended to provide a binary classification where it identifies potential students who might be at academic risk and those in good academic standing. The proposed multi-course framework aims to provide quality, early predictions of each student's academic status in weekly manner, which can help lecturers to identify students who need additional academic support and subsequently provide timely and meaningful interventions.

The framework is learning-analytics-powered, which is designed to rely solely on a series of features extracted from VLE activities and course discussion forum participations' data to provide early predictions of at-risk students. The predictive instrument is designed following ensemble modelling so as to achieve quality predictions. Furthermore, as the Grey Zone strategy described in Section 5.4.1 illustrates its ability to enhance the overall predictive models' quality and performance, we applied the same design in this chapter to optimise the model's performance. The proposed Grey Zone strategy introduces a new concept for decision-making in binary classification contexts. The design involves the implementation of base and Grey Zone models, where instances in base models' outcomes fall between pre-identified Grey Zone boundaries and are then subject to further investigation by specially designed Grey Zone models. Predictive Grey Zone models are used to re-predict underlying instances to class students as either at-risk or successful. In cases where the base model's outcome is located outside the Grey Zone boundaries, it assigns the students to one of the prediction classes directly.

Our ultimate goal is to develop a multi-course framework that is capable of providing quality predictions for future predictions across different courses settings. Hence, to examine the generalisation of the proposed framework, we evaluated the final version of model with a set of unseen testing courses. The results indicate how well the multi-course framework will act with future, brand-new samples.

6.3 Ensemble Predictive ModellingX

Ensemble modelling is widely used in a machine learning setting to develop predictive instruments (Galar et al. 2012). However, the ensemble method is used infrequently in the research field to predict each learner's performance in higher education contexts (i.e. (Boyer & Veeramachaneni 2016, Er et al. 2017)). Ensemble-based models are constructed by combining multiple classification approaches to enhance the predictive accuracy over a single learning model (Dietterich 2000). Merging several classifiers in a single model allows us to utilise a collection of hypotheses from the hypothesis space, which can help to improve prediction quality by reducing the misclassification rate.

In ensemble modelling, nominated classification members are ensembled by combining their outputs. Various mechanisms have been used to combine members' predictions to produce final prediction decisions, such as averaging members' probabilities to form the ensemble model's final output. In this chapter, we propose utilising an ensemble-based model to develop a multi-course early warning framework for at-risk students that provides quality, early predictions of students' final course achievements.

6.3.1 The Models' Development

In most cases, student performance predictive models are built by combining features extracted from single or multiple data sources, including personal and learning characteristics. This approach is suitable in cases where the object is to construct a single model where the predictive model involves the most powerful predictors from the features space.

However, we utilised an alternative methodology to develop a multi-course predictive framework by drawing together eight models, where each member model is an expert in a local area of the feature space. Local sets of features correspond to a unique type of action or analysis approach. In other words, each member model is designed to focus on predicting students who are at-risk based on a particular aspect of the students' social characteristics, textual-based analysis or a certain online learning action. Individual member models utilise features related to a single local category, where the proposed features categories are SNA, in-degree, out-degree, post weighted sentiments, post views, course module views, resources views and VLE access patterns. Figure 6.1 shows an overview of the planned ensemble model architecture.

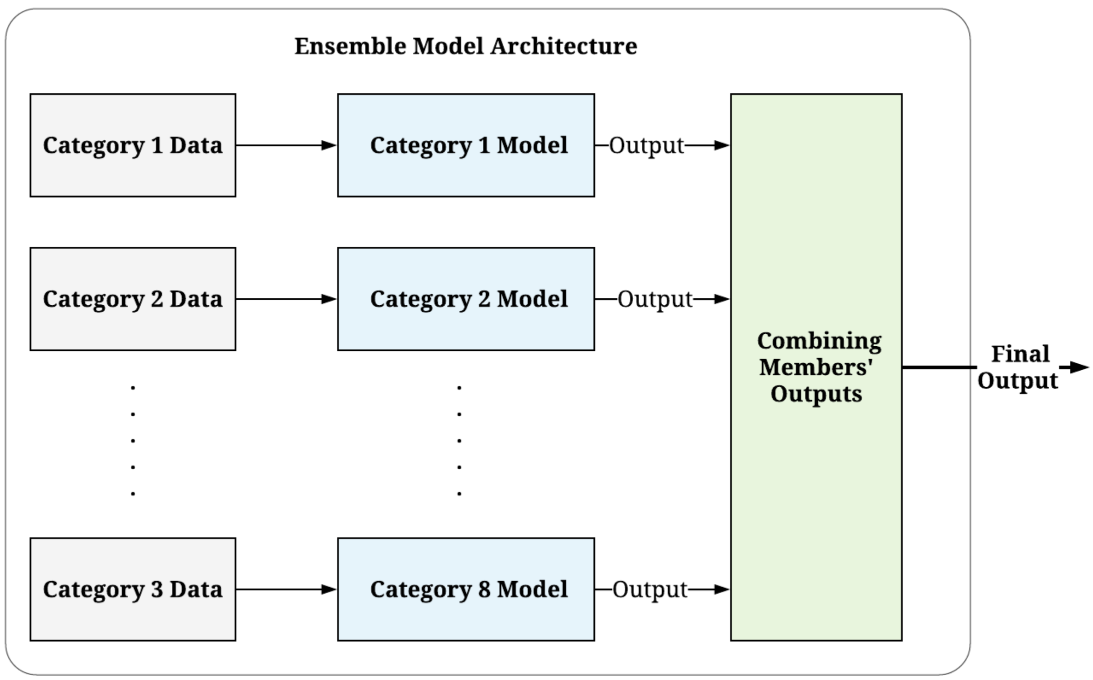


Figure 6.1: The proposed ensemble modelling architecture.

The utilised ensemble modelling design follows the divide and conquer theory, which allows us to attack the predicting students' academic risk problem by dividing it into simpler sub-problems to obtain the highest possible prediction quality. Building independent member models for each local set of students' virtual actions and social characteristics allows us to search for meaningful academic risk indicators in each subset of features, collected individually. Each member expert is constructed using the strongest subset of features within its local space of features.

In blended learning setting, online learning components are complementary to each other but with varied involvement rates. While some social learning activities are strong indicators of students' risk status, nevertheless they are rarely used by students. For instance, although features extracted from course discussion forums are decent predictors of students' risk behaviours, they are limited to those students who participate in them. On the other hand, other VLE learning components are more popular, which leads to higher interaction rates. A wide range of students engage in these learning elements, which reflect a better understanding of common aspects of students' learning performances. Thus, we take these facts into account when we select a mechanism to combine the ensemble member outputs.

6.3.2 Combining Members' Predictions

Applying a suitable outputs combination strategy is a vital element of building ensemble models, as it assigns the contribution degree of each member model on the final model output. Weighting models' outcomes is an example of the blending strategy used to combine models' outputs in ensemble modelling (Polikar 2012). Assigning weights to members allows us to manage members' degrees of importance and contributes to the final decisions. For instance, higher weightings are given to strong member classifiers, which results in higher contributions to the model's outcome.

Non-trainable and trainable weighting methods are used to allocate weight parameters associated with each single model prediction (Polikar 2006). Non-trainable combination rules allow the user to specify the weight parameters to be applied to the member models. On the other hand, a trainable weighting approach recognises the models' weights through a training algorithm, where it optimises a best-fit set of weights that produces the best performance.

The proposed ensemble model consists of eight members, where each member model is an expert in a particular category of features from the social learning global features space. Trainable weighting rules are used to optimise the ideal weighting parameters of members' models where the set of training data is used for this purpose. Then, the weighted average metric is used to compute the weighted mean of member models' predictions. While each member classifier has an associated weighting value, poorly performing members can be discarded by receiving a weight value of zero. In the weighted average mechanism, model members' outputs are combined for a given instance x using the following notation:

$$P(x) = \frac{\sum_{T}^{t=1} w_t \, M_t(x)}{\sum_{T}^{t=1} w_t} \tag{6.1}$$

Where T is the total number of model members, M_t the output vector of t^{th} member model and w_t is the assigned weighting parameter of the t^{th} member model. The key advantage of weighing model members is that it allows us to manage each members' degree of influence on the final model's output by increasing the contribution of the high-quality members and minimising the effects, or even disregarding the influence of, poor members on the model's final decisions.

6.4 Data

6.4.1 Context and Participants

Data was collected from thirteen blended computer science courses taught at the University of Adelaide, Australia, over the first and second semesters between 2012 and 2016 as described in Section 4.1. Underlying data belonging to 1,476 enrolments were collected across the courses. The utilised VLE activities datasets involves 273,983 activities logs and 3,211 posts.

6.4.2 Data Preparation

Data preparation involves a range of methods to process data, which improves its quality. Records fields in the collected datasets were validated individually. Corrupted and irrelevant logs records were either treated by filling in missing values or cleaned by removing whole records. Moreover, special attention was paid to posts' textual-content in the discussion forums dataset. While posts' textual-content were filtered from unwanted, non-ASCII characters and noise texts, the filtering process did not affect the concepts expressed in the posts. Furthermore, the collected data was clustered on a weekly basis for the length of semester. Time-series generation of VLE logs and discussion forum records were performed based on action occurrence times. Since semesters are varied in terms of academic dates, we applied a time-series clustering procedure per course based on the actual university calendar. In the clustering process, Sunday is defined as the first day of the week.

Considering the binary nature of the classification targets, students' classes (at-risk and successful) were labelled based on final course achievements. Students who earned final mark over 55 were labelled as having a successful academic status. Furthermore, students who achieved a final mark of or below 55 or who dropped out the course were labelled as at-risk. Although the course pass mark is 50, students who passed the course but achieved a final grade of or below 55 are relatively close to the failure grade.

6.4.3 Features Description

Given the restricted application of the features used in Chapter 5 to those students who choose to participate in the course discussion forum, in this chapter, we include additional online learning features extracted from VLE activities' log data to build a multi-course early warning framework of at-risk students in a blended learning setting. In Chapter 5, a range of language-based and SNA characteristics and other discussion forum aspects illustrated their ability to identify students' academic risk factors. Therefore, in this chapter, we utilised discussion forum features in addition to other features involved in VLE engagement and access patterns to cover the largest possible range of students' risk factors that can be extracted from an online learning environment.

Unlike participating in course forums, students' interactions with VLE learning components are more frequent exercises. However, each course accommodates varied types of VLE elements in terms of the nature and number of tasks, activities, and amount of available resources. Therefore, to develop a multi-course early predictive framework, we focused on key online learning elements utilised across blended courses, which are course module views, discussion forum post views and resources view activities. Moreover, online materials uploaded on the VLE may belong to either learning or assessment components. Hence, in this chapter, we merge online assignment specification views and online resources viewing actions to compute features that reflect the resources viewing activities.

A set of measures is applied to each VLE activity type to calculate features on a weekly basis throughout the semester. Extracted features cover the frequency analysis of performing a particular action, the changes in performing the action compared with the previous week and fluctuation measurements of engagement patterns throughout the semester using weekly time-series blocks of data. Furthermore, additional binary flags are associated with several features to indicate whether the resultant analysis is increasing or not. Frequency analyses are utilised to compute the appended action data from the beginning of the semester up to the current week and data is collected from actions performed in the current week only.

Moreover, the usage pattern of the VLE in terms of time exposed some aspects of students' online engagement. Several independent variables are constructed to explore daily and weekly VLE access behaviours. A student is considered connected in an individual day or week when they log into the system at least once in the specified day or week, based on the VLE logs datasets.

Finally, a preliminary analysis of each proposed feature was conducted by visualising them, which resulted in discarding several unusable predictors of at-risk students. The features are assigned to one of eight local categories of features, according to either an analysis approach or to action types. The proposed categories are SNA, in-degree, out-degree, post weighted sentiments, post views, course module views, resources views and VLE access patterns features sets. A list of features belonging to each local category is presented in Table 6.1.

Feature	**Category**
Appended degree	
Degree	
Change in degree value over 1-week	
Change in degree value over 1-week flag	
Degree trend	Social Network Analysis
Degree trend flag	
Centrality degree	
Closeness centrality	
Betweenness centrality	
Degree prestige	
Appended out-degree	Out-degree
Out-degree	
Change in out-degree value over 1-week	Out-degree
Change in out-degree value over 1-week flag	
Appended in-degree	
In-degree	In-degree
Change in in-degree value over 1-week	
Change in in-degree value over 1-week flag	
Posts' sentiment strengths	
Posts' sentiment strengths flag	

Posts' weighted sentiments

Table 6.1 continued from previous page

Feature	Category
Appended Posts' sentiment strengths	
Appended Posts' sentiment strengths flag	
Posts' sentiment strengths average	
Posts' sentiment strengths average flag	
Posts' sentiment strengths trend	
Appended course page view frequency	Course module views
Course page view frequency	
Course page views trend	
Course page views trend flag	
Change in course page view frequency over 1-week	
Change in course page views over 1-week flag	
Appended posts view frequency	Post views
Post view frequency	
Post views trend	
Post views trend flag	
Change in post views over 1-week	
Change in post views over 1-week flag	
Appended resources view frequency	Resources views
Resources view frequency	
Resources views trend	
Resources views trend flag	
Change in resources views over 1-week	
Change in resources views over 1-week flag	
Count of disconnected days in a week	VLE access patterns
First connected day in the week	

Table 6.1 continued from previous page

Feature	Category
Connected flag	

Table 6.1: A list of features belonging to each local category of the features space.

6.5 Experimental Study

The experimental study conducted in this chapter consists of four core components, including features preparation, data splitting, predictive models development and framework performance evaluation components. Figure 6.2 presents the experimental workflow for developing the multi-course early warning framework, where the process of performing each core task includes a set of sub-tasks which are described in detail in this section. The features preparation component handles the features related tasks. The experimental dataset was cleaned and the students' virtual learning features were prepared, pre-processed and assigned to the relevant categories of the eight pre-defined local categories that are presented in Table 6.1. The features preparation component involves labelling students' final achievements into binary outcomes (at-risk and successful). The data splitting component divides the full gathered dataset into training, validation and evaluation datasets. The data splitting architecture and the rationale behind utilising such an approach are described in Section 6.5.1. Training and validation datasets are used to build and tune the predictive instrument following the Grey Zone design in the predictive instrument. A detailed description of the prediction strategy is presented in Section 6.5.2. Finally, the multi-course early warning framework performance is examined in the evaluation component using an unseen evaluation dataset. Multiple evaluation metrics are utilised, including the area under the ROC curve (AUC) and confusion matrix-based measurements, including overall accuracy, recall, precision and F_1 score metrics.

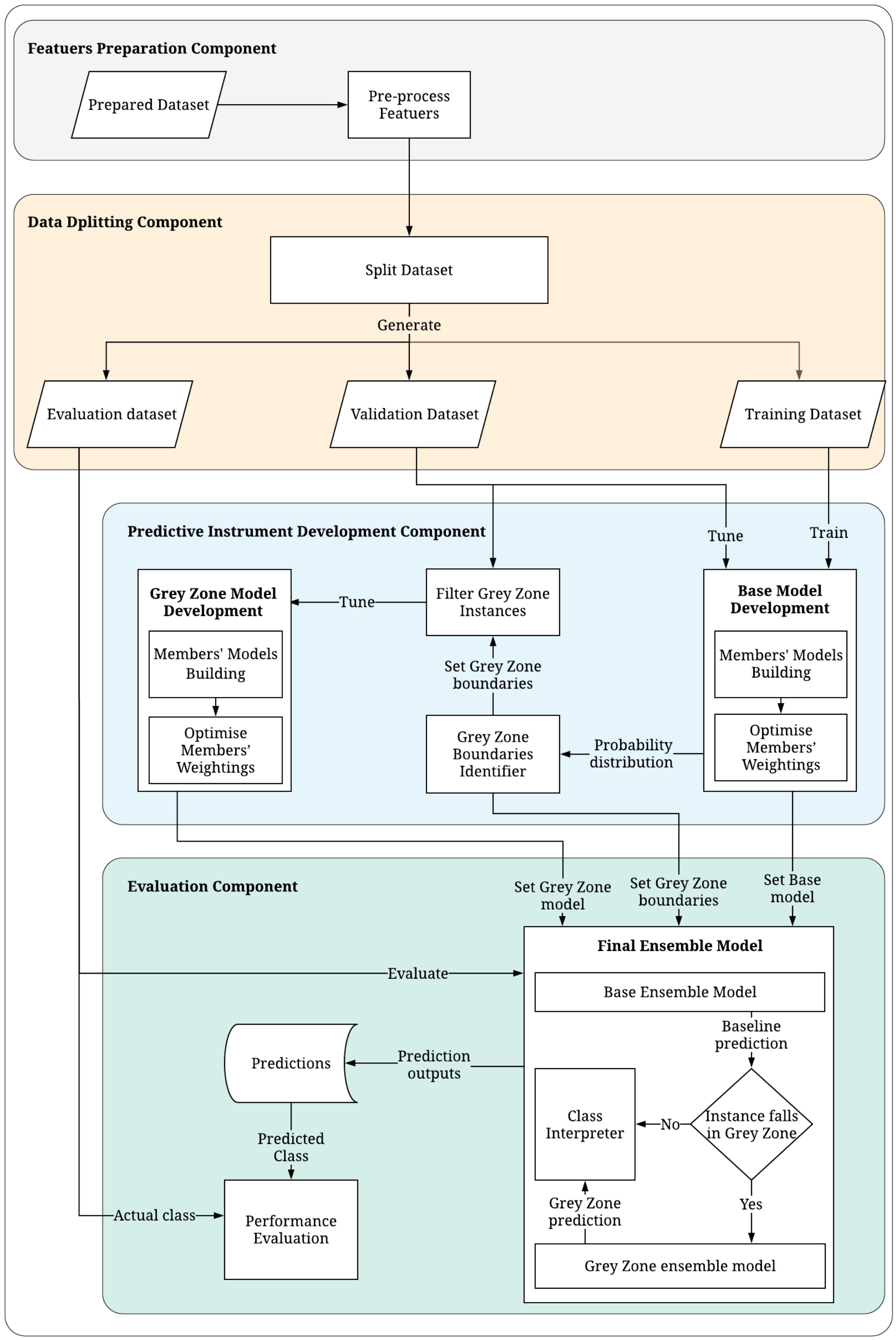


Figure 6.2: Experimental workflow.

6.5.1 Data Splitting Architecture

This thesis employs machine learning techniques to build predictive models. Therefore, prepared courses datasets were split into training, validation and testing sets to train, tune and test their predictions. The data splitting approach is presented in the data splitting component in Figure 6.2. Traditionally, the training dataset contains the largest portion of the population size since it is used as the foundation of the proposed model. A validation dataset was used during the models' development stage, where it helps to select the most influential set of features and optimise the best-fit weightings of the ensemble model members. Finally, the testing dataset contains fresh data used to evaluate the proposed model's performance and measure its accuracy. By testing the model using an unseen data set, we ensure the quality and generalisability of the predictive model and its ability to predict future cases.

A common method to split data is to perform cross-validation on a single dataset containing aggregated data from single or multiple courses. Nevertheless, in this chapter, we performed a course-based splitting approach, where all of the population belonging to the same course must be assigned to the same data category. The course-based data splitting approach was applied due to the diversity of learning structures and required effort across the courses. Consequently, by applying a course-based splitting method, we simulated real-life scenarios where students' VLE contribution levels may be different from the training and validation samples. It is important to preserve a balanced distribution of sample classes (at-risk and successful) across the split datasets to avoid undesired consequences that may arise in a model's performance or misleading evaluation results.

The original class distribution contains 53.5 percent of the population at academic risk and 46.5 percent with a successful status. An almost similar class distribution was maintained in each dataset group with some variance due to the differences in the classes' distribution across the gathered courses. The dataset splitting procedure resulted in assigning 7 entire courses (about 60 percent of the total samples) to a training set, 2 complete courses (about 18 percent of the population) to a verification set and 4 complete courses (about 22 percent of the population) to a testing set, as described in Table 6.2.

Data Class	**Course ID**	**N**
Training	Course 1	153
	Course 2	84
	Course 5	129
	Course 6	124
	Course 7	104
	Course 11	146
	Course 12	144
Validation	Course 3	132
	Course 9	136
Evaluation	Course 4	81
	Course 8	77
	Course 10	76
	Course 13	85

Table 6.2: Courses and samples sizes assigned to each split dataset.

6.5.2 Prediction Strategy

A wide variety of machine learning techniques are used to address the issue of detecting underperforming students. Logistic regression is a popular binary predictive approach in higher education research contexts. Therefore, the proposed multi-course early warning framework utilises the predictivity power of logistic regression to fulfil prediction tasks. The underlying predictive models are built using features extracted from students' discussion forum contributions and VLE interactions data clustered in weekly blocks, where the total training dataset is used to feed each member's model. Furthermore, the predictive framework utilises the Grey Zone design to improve the overall quality and accuracy of the final predictions where the design involves developing base predictive models, identifying Grey Zone boundaries and building Grey Zone models. The predictive models' construction was carried out using version 3.8.2 of the Weka machine learning library (Frank et al. 2016).

The Grey Zone concept performs by replacing the base predictive model decision by Grey Zone model predictions, which performs further investigation on instances where the base model probability falls within the Grey Zone boundaries. The Grey Zone probabilistic range is identified by performing an error analysis of the base model's initial predictions. In this chapter, the base and Grey Zone models are developed using the ensemble modelling approach described in Section 6.3.

To develop each member model, the wrapper features selection method (Kohavi & John 1997) is utilised to select the best performing subset of features from the corresponding local area of the global features space. The wrappers approach is carried out by training a set of predictive models with different combinations and subsets of features, followed by a comparison of the models' performances. In the utilised ensemble modelling design, each member is examined using pre-defined sets of features that are expert in a particular set of online learning engagement characteristics. The weighted average metric is used to combine members' predictions and define the format of the final model's outcomes. In this chapter, AUC and overall accuracy metrics are used as criteria to select the best performing member models, as well as using the best-fit weighting parameters with the help of the validation dataset. Finally, the base and Grey Zone models are combined into a single predictive instrument that presents the predictive component of the multi-course early warning framework. The developed framework evaluation is performed using an unseen dataset that combines data drawn from varied courses. The rest of this section describes the development component of the predictive model.

Building the Base Model

In the Grey Zone strategy, the base models generate initial predictions where the strategy aims to address the weakness of such models by identifying a probabilistic range where most of the prediction errors occur and replaces these predictions using alternative models. Therefore, the quality of the base models is an essential element of the overall predictive instrument's performance.

The proposed framework aims to predict academic risk patterns on a weekly manner across the core study weeks using online learning activities. Hence, the framework utilised features that were grouped into eight categories, where each category corresponded with learning or social characteristics. Each base ensemble model member is implemented strictly using the best performing sub-set of features falling under a single local category, where the predictive models are fed with the semester's aggregated dataset. Subsequently, we built the eight model's members, we ran the members' weighting script to optimise the best-fit weighting parameters for each member.

The members' weighting task is carried out by applying different collections of weighting values followed by comparison of the resultant predictions' quality and accuracy. Features selection and optimising weighting tasks were completed with the help of a verification dataset, where it was used to examine the performance of different combinations of features' and outputs' weighting parameters across all the lecture weeks. Finally, different combinations of members' models and weighting parameters were evaluated to select the best-performing ensemble base model, which would provide the most reliable predictions across all the prediction weeks.

Identifying Grey Zone Boundaries

A key part of the Grey Zone design is identifying the weaknesses in the base model which are then identified as a Grey Zone in terms of a probabilistic range. In other words, the Grey Zone covers the range of probabilities, computed by the base model, which have a high misclassification rate. Therefore, we performed an error analysis of the base model estimations to indicate where the Grey Zone upper and lower boundaries fall. Since a single base model is used for all of the prediction weeks, verification instances were aggregated into a single dataset to identify a single set of boundaries for all the study weeks.

Firstly, we visualised the distribution of the semester-aggregated base model probabilities, highlighting where the range of probabilities in the base model outcomes overlapped in terms of each prediction class, across all the weeks. Then, with the help of an ROC graph, we evaluated the base model's performance at each possible threshold point in the prediction space to identify the best Grey Zone cut-off values. Finally, the resultant collection of Grey Zone boundaries was analysed to select the optimal pair boundaries parameters.

The Grey Zone boundaries were selected based on the model's initial threshold value of 0.5. The nominated cut-off value of 0.55 delineated the upper boundary and the value of 0.45 was made the lower boundary of the Grey Zone across all prediction weeks. The proposed Grey Zone has a noticeably high misclassification rate, where about 73 percent of the base model's misclassifications occur within its boundaries over the 12-study weeks.

The base model performs relatively well outside the identified Grey Zone boundaries. To address the high error rate in the Grey Zone, instances that have base model estimations that fall within these boundaries are subject to being re-predicted using the weekly Grey Zone models.

Grey Zone Models

Grey Zone models are designed to complement the base model's efficiency, where they replace the base model when it performs poorly. Therefore, students who fall in the Grey Zone are subject to being re-predicted via the Grey Zone models. Grey Zone models are built specifically to distinguish risk factors of instances that fall in the Grey Zone. Hence, only corresponding Grey Zone instances are used to validate Grey Zone model components for each underlying prediction week.

Since the extracted online learning characteristics for Grey Zone instances change significantly each week, a Grey Zone model is developed specially for each study week utilising the same ensemble modelling design as was used to develop the base model. However, as the mid-semester break starting week is varied every semester, a special Grey Zone model is implemented for the week prior, regardless of where the week falls in the actual study sequence. The development of a Grey Zone pre-break model is motivated by the dramatic change in students' behaviour in this study week. Therefore, the set of Grey Zone models involves 13 different Grey Zone models, 12 models corresponding to core study weeks and a model designed especially for the week prior to the study week.

The Grey Zone models' development process followed the same procedure of building a base model to implement the Grey Zone models. However, only that specific week's Grey Zone instances were used for feature selections and weighting optimisation tasks, which enabled us to focus on the risk aspects of the utilised subset of instances.

Every Grey Zone ensemble model contains eight member models, where each member model is an expert in the local area's features space. Members' models are built using the best performing bag of features within the related features list. After developing the member models, the members' output weighting task was performed to optimise the best-fit weighting parameters for each member. Then, the performance of each Grey Zone ensemble model was examined using AUC and an overall accuracy metric to ensure its performance against the base model's predictions. The resultant Grey Zone models utilised different sets of features and weighting values across each model, which can be explained by the change in the features' importance across each study period.

6.6 Results

This chapter aims to develop an exemplar multi-course early warning framework for at-risk students based of online learning data following the Grey Zone design. To assess the developed multi-course early warning framework's capacity to be generalised and predict future data, we performed evaluation task using four whole, unseen courses. Although all four testing courses were taught at the School of Computer Science at a single institute, each course is structurally different and contains varied distributions of student involvement. Therefore, the virtual learning pattern is diverse in terms of the types of activities and frequency of performing online actions. Heterogenous evaluation data allows us to evaluate the multi-course framework under various real-time scenarios in order to provide a reliable assessment of its predictive quality.

Figure 6.3 presents a spider chart of the overall virtual learning activities distribution across the evaluated courses, which highlights the differences in the courses' online learning properties across evaluation courses. It is generally recognisable that the popularity of the VLE login action as it presents gateway to navigate through the VLE components. Furthermore, post views present as the second most popular online learning activity. However, the viewing posts action is extremely affected by the number and quality of messages posted on the course discussion forum. In the same way, resources view actions are also linked to the type and quantities of online materials available in the VLE environment. For instance, evaluation course 3 has the highest number of learning resources available online. Moreover, in evaluation course 1, the online resources are limited to assignment specifications, which explains the low number of interactions with the resources components. Moreover, posting messages on the course forum is the lowest performing activity across all evaluation courses.

Before evaluating the results, it is important to rank the importance of the evaluation metrics. Given the binary nature of the problem and the fact that the evaluation

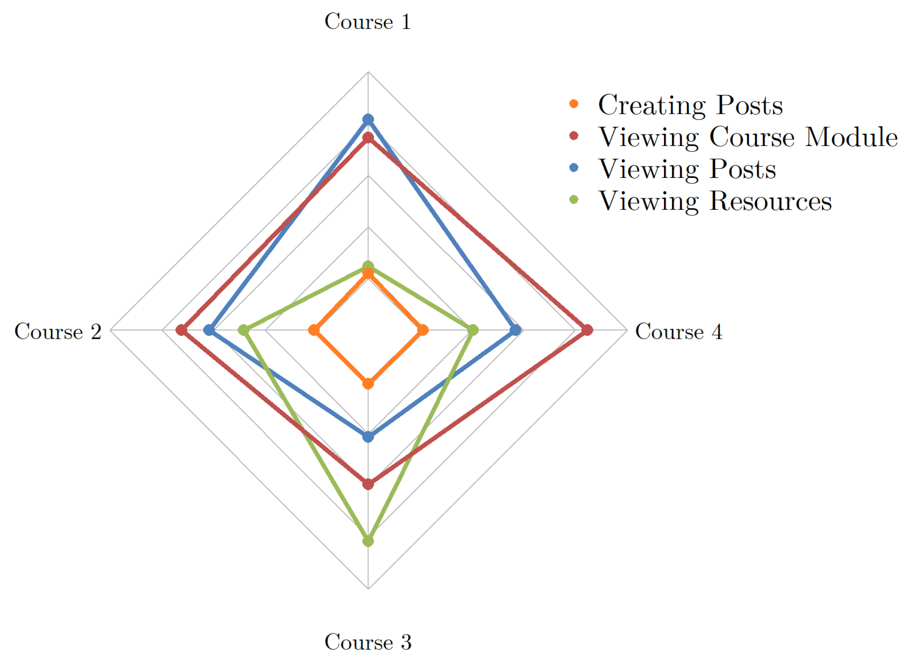


Figure 6.3: A spider chart of students' virtual learning activities' distribution across the evaluation courses.

courses might have diverse distributions of classes, we utilised AUC metrics to examine the quality of the predictions, as well as the overall accuracy of the framework predictions. Furthermore, we also employed other typically used confusion-matrix-based measures, including overall accuracy, recall, precision and F_1 score metrics. Two types of error might rise throughout the evaluation process: at-risk students who are misclassified as successful mistakenly (false negatives) and successful students who misclassified as at-risk mistakenly (false positives).

The developed automated multi-course early warning framework of at-risk students aims to help to improve academic outcomes by identifying at-risk students, so that interventions can be made for students who need additional academic support. Therefore, the predictive instrument must be designed to provide the highest possible number of reliable predictions for both classes (at-risk and successful) in the early stage of the semester. This is motivated by the issues that are associated with giving one class a higher weighting than the other in the prediction space, which may lead to delays in providing proactive support for at-risk students or even not to provide it at all. On the other hand, identifying students in good standing as at-risk can lead to unnecessary workload for lecturers.

A series of experiments have been carried out to examine the robustness of the proposed multi-course early warning framework when it predicts unseen datasets. Each evaluation course was tested individually throughout the 12 study weeks to evaluate the framework's predictive performance on each testing course separately. Table 6.3 presents the experimental results of each single evaluation course dataset in terms of its AUC and accuracy metrics. Furthermore, Figure 6.4 visualises the model's weekly evaluation performance, using multiple evaluation metrics for same evaluation dataset.

Generally, the experimental results show that the multi-course predictive framework was able to produce high-quality predictions across three evaluation courses, while performing decently in one evaluation course. Although the framework's predictivity was relatively low in the first prediction week, there is a noticeable improvement in the framework's predictive power as the semester moves towards the end for all evaluation courses. The predictive framework performance rises due to the increase in virtual learning tracing data that become available each week. However, the framework's accuracy dropped in the final prediction week due to the dramatic decrease in students' online learning activities across all evaluation courses.

With regards to overall performance evaluation, the multi-course predictive framework generates accurate predictions across most of the evaluation courses with similar

	Course 1		Course 2		Course 3		Course 4	
	AUC	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**	**AUC**	**Acc**
Week 1	0.67	58.0 %	0.62	58.4%	0.68	63.2%	0.76	65.9%
Week 2	0.62	56.8%	0.73	68.8%	0.83	72.4%	0.9	78.8%
Week 3	0.66	64.2%	0.89	80.5%	0.85	71.1%	0.86	82.4%
Week 4	0.64	63.0%	0.88	84.4%	0.84	75.0%	0.88	82.4%
Week 5	0.59	63.0%	0.9	88.3%	0.86	80.3%	0.85	83.5%
Week 6	0.65	60.5%	0.81	83.1%	0.83	77.6%	0.86	81.2%
Week 7	0.56	63.0%	0.9	85.7%	0.9	84.2%	0.91	84.7%
Week 8	0.72	76.5%	0.88	80.5%	0.86	81.5%	0.9	85.9%
Week 9	0.81	77.8%	0.94	89.6%	0.88	82.9%	0.92	87.1%
Week 10	0.76	76.5%	0.93	83.1%	0.88	81.6%	0.92	85.9%
Week 11	0.72	76.5%	0.86	88.3%	0.93	89.5%	0.84	84.7%
Week 12	0.77	77.8%	0.89	83.1%	0.86	79.0%	0.9	87.1%

Table 6.3: A demonstration of the weekly prediction performance on four fresh evaluation courses.

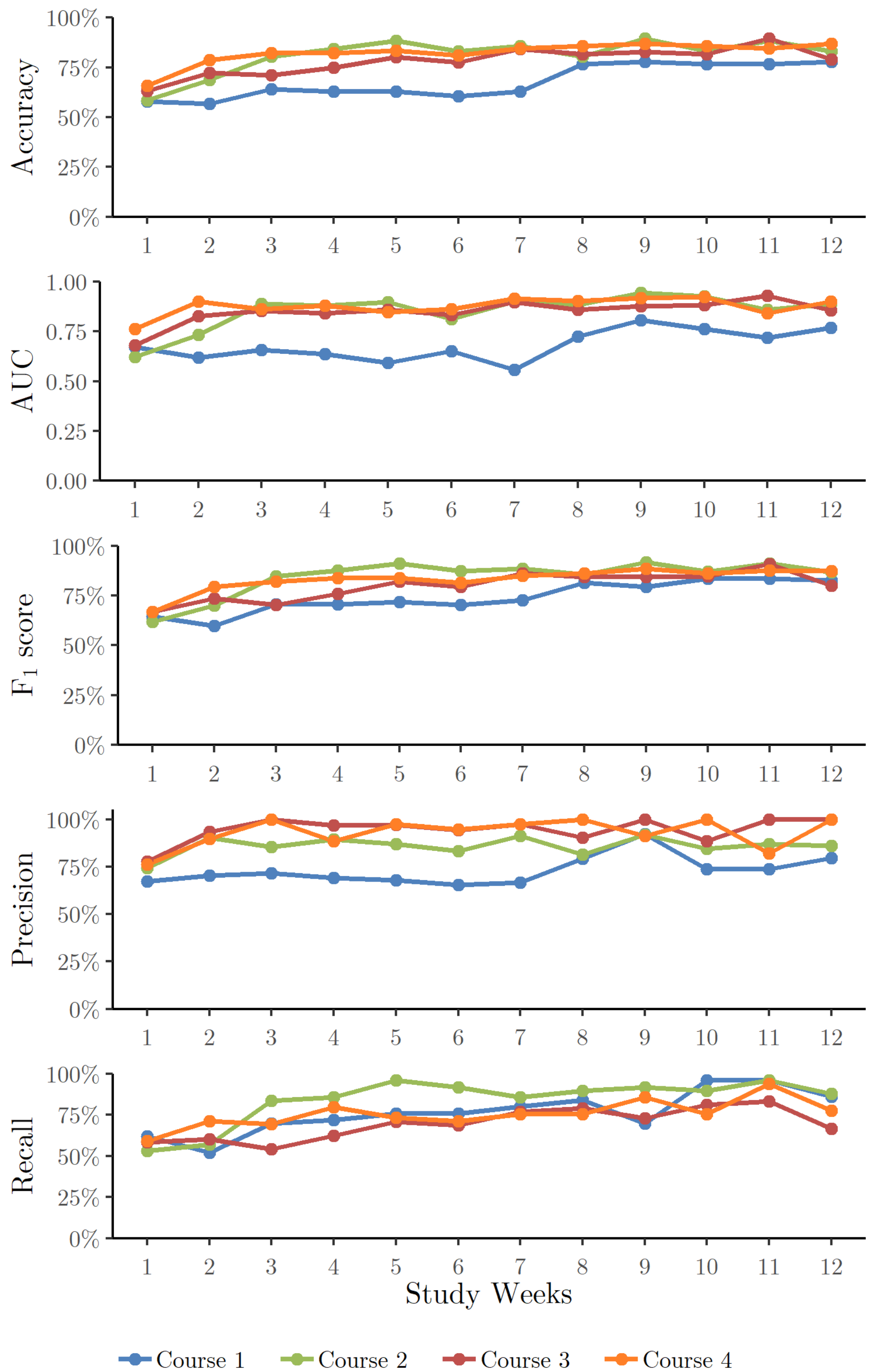


Figure 6.4: Plots of the evaluation dataset evaluations in terms of AUC, accuracy, F_1 score, precision and recall metrics.

improvement patterns in prediction quality throughout the prediction weeks in terms of both AUC and accuracy metrics. While the predictive framework's quality mostly continued to improve in terms of the AUC metric, its performance, as measured by the overall accuracy metric, dropped in prediction week 12 by 3 percent on average. When it comes to evaluation course 1, the predictive framework produced a lower prediction quality when compared with the other evaluation courses in the early prediction weeks, but it increased in performance significantly after study week 7.

The framework's predictive performance is varied across the evaluation courses. The multi-course predictive framework achieved over 0.92 AUC points across most of the evaluation courses, while obtaining top AUC value of 0.81 in one course (Course 1). With respect to its overall accuracy, most of the evaluation courses reached decent prediction accuracies by obtaining an accuracy over 87 percent.

Furthermore, the evaluation results indicate that the predictive framework's ability to detect at-risk students was over 93 percent across most of the evaluation courses, with an ability to avoid misidentifying any successful student with a precision value of 100 percent in numerous cases. However, the recall measures fluctuate throughout the prediction period, while the precision movement tends to be stable. The harmonic measure between precision and recall shows the framework's capability to achieve accurate classifications in both classes (a-risk and successful). The framework achieved an F_1 score value of 80 percent on average across all evaluation courses throughout the semester, reaching a maximum harmonic value of 92 percent.

Overall, the evaluation results confirm the ability of the developed multi-course early warning framework to produce quality predictions of students' academic risk status using online learning data across multiple heterogeneous courses.

Despite the fact that the multi-course framework improves its predictive quality in most of the prediction weeks as more data become available, the framework produces consistent predictions in instances drawn from evaluation course 1 with a very slight improvement for more than the half the semester due to the low VLE participation rate in the course. For instance, evaluation course 1 has significantly lower resource view actions compared with other evaluation courses due to the very limited number of resources uploaded onto the course online module. Nevertheless, in terms of the underlying evaluation courses, its predictive performance started to rise after study week 7.

Then we combined the weekly testing courses results to determine the optimal time when lecturers should deliver academic support to students in-need. We conducted an error analysis of the combined results to identify the ideal week where course lecturers should deliver interventions to students who have been identified as at-risk academically. Considering the necessity for providing early feedback and corrective actions to achieve the best possible outcome and the framework's ability to start producing reliable predictions, we consider week 3 as the optimal week to establish providing additional support to at-risk students. At week 3, the early warning framework was able to achieve quality predictions over 0.8 AUC point in the aggregated testing dataset. Figure 6.5 shows the predictive framework's weekly performance across all evaluation courses in terms of the AUC metrics and identifies the optimal time to provide interventions to students in need of additional academic support.

Finally, although the predictive framework developed in this chapter illustrates its ability to produce quality predictions as early as study week 3 across multiple unseen evaluation courses, it still contains some limitations caused by a range of pedagogical aspects. The multi-course early warning framework is powered by students' VLE interactions data, which is affected by course characteristics such as the specific VLE module design, amount of learning resources made available online for each week and the lecturers' level of collaboration with the students in the course. Another limitation is related to identifying student's personal weakness aspects in the learning process, which, if solved, would help instructors to plan personalised intervention actions to suit each at-risk student.

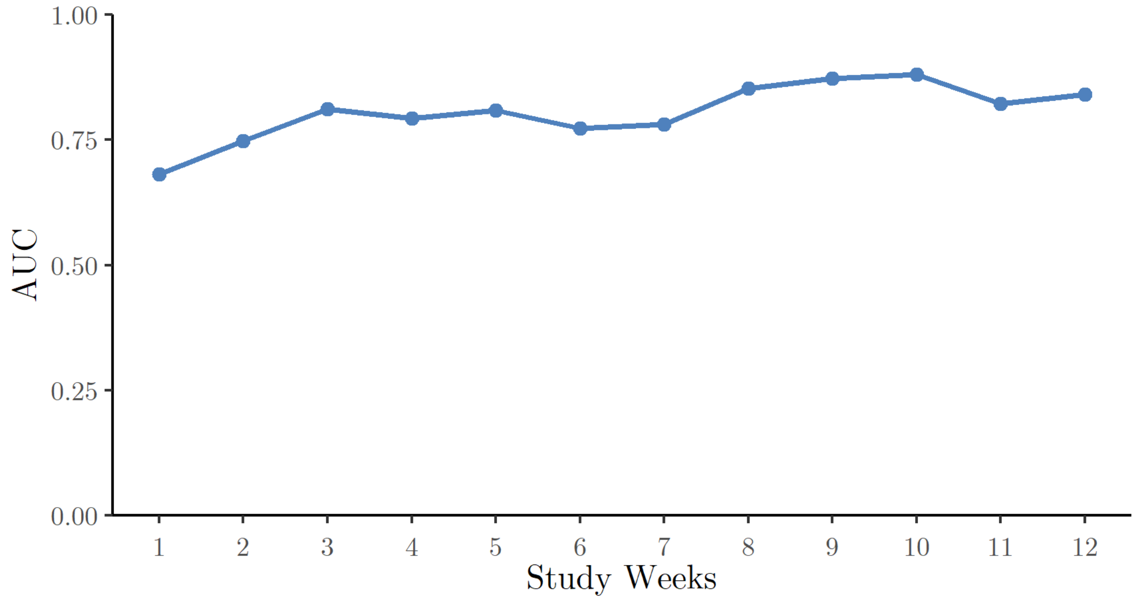


Figure 6.5: Weekly predictive framework performance based on aggregated evaluation dataset results in terms of the AUC metric.

6.7 Summary

In this chapter, we developed an exemplar multi-course early warning framework that detects students who might be at failure or attrition risk, in a weekly manner, with decent overall performance. The framework was built using an ensemble modelling strategy, where the ensemble predictive model consists of eight weighted members; each individual member is developed using a unique set of features belonging to a single category. Each features category combines a unique subset of features from the global space. Furthermore, the multi-course early warning framework was implemented with the help of a Grey Zone decision-making strategy to improve the prediction accuracy by applying additional investigations to instances falling within the boundaries of the Grey Zone, where a high proportion of the misclassifications occur.

The framework is evaluated with using four entire unseen courses datasets (N = 319 enrolments) to examine its ability to predict future instances where evaluation courses have a variety of VLE activities distributions. The predictive framework's top performance ranged from 0.81 to 0.94 AUC points across the evaluation courses. With regards to accuracy metrics, the framework obtained its best performance, between 77 and 90 percent, across individual courses in the evaluation dataset. The framework was able to provide reliable predictions as early as week 3 of the semester, when early interventions can be provided to support students in need.

Moreover, while the multi-course early warning framework developed in this chapter illustrates its ability to predict future events with high classification rates across the majority of the evaluation courses, its predictivity power is limited to students' learning characteristics included in the training concepts. This limitation reduces the framework's ability to distinguish the changes in students' learning characteristics in different educational settings. Therefore, in the next chapter (Chapter 7), we extend our work by developing an adaptive learning mechanism to accommodate additional online learning patterns, which may be observed in a newly obtained dataset to improve the framework's performance and ability to provide accurate and early identification of at-risk students.

Chapter 7 applies an adaptive in-system analytics approach to update the framework's predictive instrument and its properties as soon as a new dataset becomes available. The proposed adaptive approach observes and adapts to new learning patterns to enhance the framework's predictivity. In the next chapter, we aim to make the framework able to learn from its previous predictions, automatically. The framework will be fed with extra student VLE interactions and discussion forum datasets associated with final achievements to train and validate the framework's predicative instrument to consider outcoming patterns.

Chapter 7

Towards an Adaptive Early Warning Framework for At-Risk Students

7.1 Overview

Every day, a huge volume of data is generated from students' engagements with various VLE components. While VLEs are mainly used as learning tools, students' digital footprints reflect their learning progress, which can be used to predict their final course outcomes. Many attempts have been made to analyse students' online learning traces alongside other data sources. The majority of the existing efforts rely on machine learning algorithms to learn students' personal, academic and learning characteristics to forecast course outcomes. However, existing predictive models in the literature are devoted to static machine learning environments, where models tend to be trained on historical information and remain fixed with no updates (i.e. (Bainbridge et al. 2015, Cen et al. 2016, Chai & Gibson 2015, Howard et al. 2018, Hu et al. 2014, Jishan et al. 2015, Natek & Zwilling 2014)). The assumption behind this approach is that there are no significant variations in learning patterns across different learning environments or that VLE usage patterns would not change over the duration of the predictive model's use.

On the other hand, some concerns have arisen regarding the scalability of such predictive instruments, as students' learning behaviours might vary due to cultural or environmental reasons (Leitner et al. 2017). These concerns are extended to include a predictive models' ability to cope with changes that may occur in students' learning styles over time. These situations increase the need to develop and integrate an adaptive mechanism, allowing student performance predictive models to adapt to new cultural or educational settings dynamically to enhance the quality of the models' predictions. To update the predictive instruments, adaptive approaches are powered by brand new data gained from fresh cultural or educational settings, or by using in-suite up-to-date datasets which reflect the current status of data distribution.

Concept drift refers to the problem when users change their interaction patterns, which results in changes in input data or the predictive instrument's outputs distributions (Gama et al. 2014). While the problem has been studied widely in many branches of machine learning and data mining contexts, this problem has not been studied, to the best of this author's knowledge, in terms of predicting students' academic performance. Drift may occur in students' learning behaviours due to several reasons, such as changes in students' learning patterns, students' learning preferences or course structures. In these situations, adaptive strategies subject the predictive instrument to constant change in its properties to accommodate new learning patterns or adapt to new prediction settings via a single or a combination of sets of adaptive learning mechanisms.

Adaptive learning methods include a set of approaches to update predictive models to optimise their performance by utilising adaptive mechanisms such as re-training or updating existing classification models using recently obtained data batches, which enhance the training data coverage or use newly obtained datasets to restructure underlying predictive models. Other adaptive mechanisms are designed particularly for ensemble learning, including dynamic inclusion of member models, replacing existing member models by new, better performing predictive models and re-optimising the outputs' combinations of parameters. Adaptive approaches aim to enhance the predictive models' ability to update their parameters according to the latest learning patterns, which can improve their predictive power. However, performing the adaptive process manually can be a time-consuming process, particularly with the heavy workload that lecturers faces in the contemporary educational system.

Therefore, this chapter introduces an adaptive learning framework capable of adapting to changes in students' learning behaviours or learning environments. The proposed multi-course, early warning framework is relevant to the Grey Zone design proposed in this study. Furthermore, as the Grey Zone modelling combines base and Grey Zone models in one model, it is convenient for updating all the predictive components, alongside the Grey Zone configurations. In terms of adaptive mechanisms, we propose deploying a combination of multiple adaptive approaches to adapt to the potential changes in data distribution that may arise in adaption data batches. The adaptive solution is designed to handle updates in predictive ensemble learning in conjunction with the Grey Zone strategy. The solution involves two algorithms: the first algorithm is designed to update the base and Grey Zone predictive models, while the second identifies optimal Grey Zone boundaries. To the best of our knowledge, this is the first work introducing adaptive learning concepts into the prediction of students' performance context.

The rest of this chapter is organised as follows: the chapter starts by shedding light on popular adaptive mechanisms used in the literature to update predictive ensemble models in Section 7.2. Moreover, this section discusses possible adaptive methods applicable to the Grey Zone design. In Section 7.3, we describe the proposed Adaptive Grey Zone Ensemble Model (AGZEM) framework. Subsequently, we describe details of the Ensemble Model Adaption (EMA) algorithm and the Grey Zone Bounds Adjustment (GZBA) algorithm developed for this work. The proposed adaptive framework and algorithms are the vehicle for the experimental study conducted in this chapter, which is presented in Section 7.4. This section presents an experiential dataset, setup and results, which evaluate the effects of the proposed adaption solution in the context of predicting students' academic performance. The experimental results suggest that the adaptive solution improves the overall quality model performance in terms of the AUC metric. The adaptive approach is powered by coping with changes in recent data batches, which might vary from the current concept. Therefore, the results also suggest that utilising a forgetting mechanism for irrelevant data instances may help to achieve faster and better adaptation outcomes. Finally, we finish this chapter with a brief summary.

7.2 Adaptive Approaches

Generally, predictive instruments are required to cope adaptively with unexpected changes in users' behaviours and environmental settings as they occur over time. The instruments' ability to update their structures and properties by incorporating recent data may maintain their capacity to provide quality predictions and improve their predictive power. A fresh dataset involves the most advanced status of data distribution. Therefore, it is widely assumed that more recent data have higher relevance to the concept than historical instances. Hence, a higher importance is usually given to newer examples throughout the process of updating predictive models.

In machine learning contexts, various adaptive learning algorithms have been developed to enhance predictive models' ability to deliver more accurate estimations. These algorithms employ different adoption methods. Numerous adaptive mechanisms are concerned with updating individual predictive models, while others focus on updating or adjusting ensemble models. In this section, we present popular adaptive approaches discussed in the literature that are relevant to adapting ensemble learning models alongside suggesting adaptive strategies applicable to the Grey Zone design. Ensemble learner adaptive strategies are discussed in terms of the methods can be used at individual member level, adapting ensemble model structures and adapting members' outcome aggregating methods. Though several adaptive learning algorithms contain a single adaptive mechanism to update predictive models, others utilise hybrid approaches.

Moreover, a related issue regarding adaptive methods concerns the controlling factors used to execute adaptive processes. Various triggering conditions have been described in the literature. Numerous adaptive strategies rely on instruments to detect changes in the data, and then, whenever changes are detected, the adaptive process is activated (i.e. (Chu & Zaniolo 2004, Gama et al. 2004, Wang et al. 2006)). Moreover, a diverse collection of mechanisms has been used to detect changes in the underlying data distribution, such as tracing the changes in the outcome probabilities density (Gama et al. 2004, Wang et al. 2006). In another case, Alippi et al. (2012) attempt to assess changes in input data using two different mechanisms based on sample distributions and prediction errors to trigger the classifier reconfiguring process. Furthermore, other adaptive frameworks utilise more regular approaches to control the adaptive process, where it is subject to a stream of data batches (Raza et al. 2015). However, several adaptive methods do not employ event-related triggers, as they activate the adoption process at routine timeframes (i.e. (Shalizi et al. 2011)).

The rest of this section discusses widely used adaptive mechanisms to update machine learning ensemble learners. Mechanisms are performed on different levels, including updating ensemble model members, changing predictive models' structures and adjusting output combinational parameters. The section also presents a number of adaptive methods that can be used to update Grey Zone modelling configurations dynamically.

7.2.1 Training Adaptivity

Machine learning involves a wide range of learning algorithms that can be used to build members' predictive models in ensemble modelling. In predicting student academic performance settings, various learning approaches were gathered together to build predictive models, such as mixing a decision tree with AdaBoost approaches in one model and blending CART with an AdaBoost approach to create another predictive model (Hu et al. 2014), and blending a decision tree, gradient boosting, rule induction and regression in (Shelton et al. 2016, 2017).

Many mechanisms have been used to update ensemble model members, including training members using an increment training dataset. Feeding the ensemble model's members with an updated training dataset is an effective approach for adopting additional observations to predictive models. Newly arrived data instances increase the coverage of the training dataset, which may reflect positively on the predictive model's performance. However, predictive techniques have a different nature with respect to their adaptation ability. While some classification methods can handle newly obtained training datasets implicitly (e.g. k-nearest neighbours (Hastie et al. 2009)), others are designed for static datasets, such as logistic regression. In the latter approach, models must be updated explicitly whereby the models are retrained on appended training datasets. Moreover, an additional adaptive approach is related to utilising dynamic data pre-processing methods to enhance data quality and consequently member models' performances (Zliobaite & Gabrys 2014).

On the other hand, an historical training dataset may contain an irrelevant concept. Therefore, some adaptive mechanisms consider discarding such irrelevant data instances using forgetting strategies to minimise the impact of old concepts on updated model decisions. Sliding windows and decay factors are popular forgetting approaches for old training datasets.

7.2.2 Structural Adaptivity

Restructuring ensemble models is a well-known adaptive methodology to enhance global model performances, instead of focusing on each member's performance. On a member model's level, the updating process can be performed by applying various methods, particularly training-related and structural-related approaches. A structural adaptive strategy can be applied by updating the set of utilised features to suit new changes, employ new features, or even to employ a different learning algorithm.

On the ensemble learner level, structural adaptive strategies involve adding additional members or replacing, removing, activating, or deactivating existing ones, to enhance the underlying ensemble learner's flexibility to adapt to changes in the concept environment. Adding fresh well-performed models fed with an updated training dataset or exchanging underperforming models with a better performing model helps to refine global performance. Furthermore, activating existing member models after deactivation may enhance the overall ensemble model's performance.

Adding, replacing, and activating predictor members is usually made as a response to changes in the model's predictive performance. Various strategies have been applied to monitor ensemble learner quality and trigger the adaptive mechanism to add, replace or activate members. For instance, Kolter & Maloof (2003) proposed an approach to handle changes in data distributions, whereas the approach adds a new model member after each classification mistake occurs with allocated weight of 1. Members' weighting parameters are adjusted dynamically: thus, where the parameter reaches a value below a pre-defined threshold, the corresponding member is removed. However, such an adaptive approach may lead to a significant increase in the amount of ensemble model members. Another technique handles updating ensemble learners by training new classifier on new data instances, then the newly-built model replaces the worst-performing existing member (Street & Kim 2001). Other studies have proposed other mechanisms to control updating ensemble models, such as strategies that rely on obtaining new examples (Raza et al. 2015) and others again retain a set of deactivated model members trained on old datasets to be re-activated in case an ancient concept repeats (Soares & Araujo 2015).

Removing or deactivating redundant members is a common adaptive approach used to update ensemble learners, which results in excluding insufficiently performing members (Bouchachia 2011). Discarding member models with insufficient performance results in favour of the more significant influence of quality members on the final prediction outcome will lead to better overall predictive performance (Soares & Araujo 2015). An alternative adaptive mechanism associated with ensemble modelling is applying weights to each member model's outcomes, where under-performing members receive low or null weightings. While a low contribution weight results in limiting the effect of members' final outputs, weighting the value of zero means that corresponding member models have been deactivated. Further explanations of such weighting adaptive approaches are discussed in the following section.

7.2.3 Combinational Adaptivity

A viable alternative adaptive method for ensemble learners is modifying the combination parameters. The ensemble learner obtains its final estimations by combining multiple members' predictions. The outputs combination method can be designed to react dynamically to changes in incoming data. Commonly, ensemble learners employ combination schema, weighting parameters associated with each individual member (Polikar 2012) which rule individual member's contributions to the final decision. Weighted ensemble learners typically utilise an updatable combination design where the model's members can be adaptively reweighted to respond to changes in recent data distributions.

Several weighting criteria have been used in order to fine-tune members' weights in an adaptive ensemble learning setting. For instance, in numerous adaptive frameworks, members are weighted based on their performance after being trained on sequential batches of data, such as in (Wang et al. 2003). Furthermore, others consider various factors during the adaptive weighting process, such as members' ages and examining changes in error rates (Elwell & Polikar 2011).

The process of reweighting members overlaps with structural adaptive methods by having the ability to disable the contribution of single or multiple members, for instance, in the case where a weighting value of zero has been assigned to a weighting parameter, which means the corresponding member is deactivated by having null influence on the final model's outcome. In the same way, when a very low weight is given to an individual member, that reduces its ability to influence the model's decision.

7.2.4 Configurational Adaptivity

In this section, we suggest adaptive mechanisms that can be associated with the Grey Zone design. The Grey Zone concept works by identifying a probabilistic range where the majority of misclassifications occurs in the base model and replaces the base model prediction by a better-performing Grey Zone model for instances that fall in the Grey Zone. Besides the adaptive methods mentioned earlier, several approaches can be developed to re-configure the structure of the Grey Zone model.

A core element of the proposed design is identifying Grey Zone boundaries that are consistent with the base model. The design can cope with changes in the base model output distribution by re-configuring the upper and lower boundaries adaptively, based on either recently obtained batches of instances, or according to the updated base model properties. Furthermore, fitting an additional Grey Zone model dynamically to handle special events might be observed in a newly obtained dataset. For instance, in the context of predicting students' performance, a need for a new Grey Zone model may arise to handle if a study week falls immediately after the mid-semester break, which can be constructed and fitted adaptively. Moreover, the design can cope with changes in prediction concept by observing the performances of Grey Zone models whenever a change in data or the base model occurs and handle new concept adaptively either by updating or replacing the existing Grey Zone models.

7.3 Adaptive Grey Zone Ensemble Model Framework

In this section, we introduce the proposed Adaptive Grey Zone Model (AGZEM) framework for data-driven continuous improvement. The AGZEM framework is specifically designed to comply with the proposed Grey Zone strategy characteristics and ensemble modelling design used to develop the multi-course early warning framework. Then, we present the adaptive algorithm proposed to cope with recently obtained batches of data.

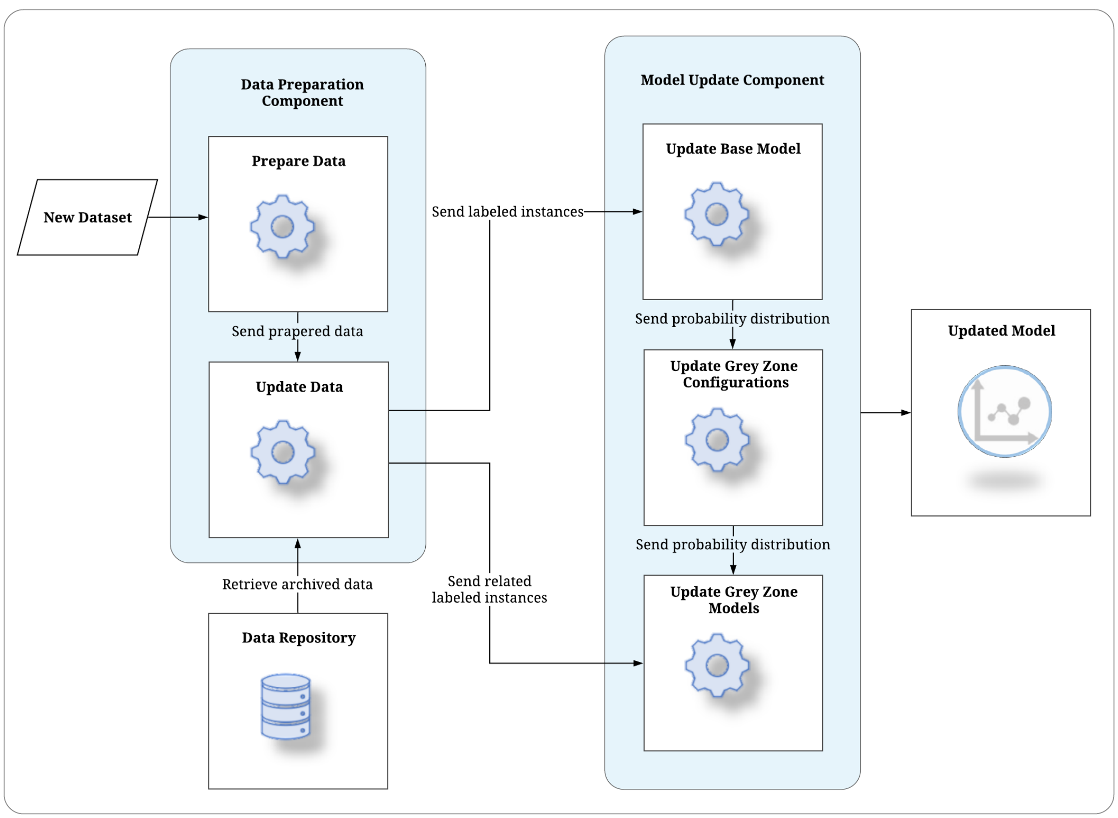


Figure 7.1: The high-level architecture of the AGZEM framework.

7.3.1 AGZEM Framework Architecture

Figure 7.1 depicts the high-level architecture of the AGZEM framework. The proposed architecture consists of two key components, namely the data processing component and the predictive models' update component. In this study, we suggest triggering the adaptive framework based on data streaming factors as data is the major driver of the adoption process. In other words, the adaptive process is activated whenever a batch of new data is fed into the framework.

Data processing component: The data processing component handles several data tasks, which involve data preparation mechanisms, updating the stored dataset and providing an adaptive algorithm with appropriate training and testing datasets. Moreover, this component is responsible for interacting with the data repository, where archived data are maintained. The data preparation module includes multiple tasks to prepare the recently obtained dataset for analysis, particularly features extractions, encoding target labels, pre-processing and clustering observations on a weekly basis, to match the exact same format as for the historical dataset. Furthermore, updating the dataset module responsible for integrating the retrieved historical observations with the new ones in a single dataset. The last element controls the data splitting mechanism to supply the adaptive component with appropriate training and testing datasets to perform the update process. For instance, in the case of updating Grey Zone models, the testing dataset is limited to instances that have a corresponding base model outcome, falling within the Grey Zone boundaries for underlying sets of weeks or events.

Predictive model's update component: the predictive model's update components are responsible for updating all the predictive models associated with the ensemble learner and their parameters. Since the underlying ensemble model follows the Grey Zone strategy, it groups two types of models; the base predictive model and the Grey Zone predictive models. The ensemble learner involves Grey Zone properties, which may need to be reconfigured. Therefore, the adaptive process involves three stages where the base model must be updated first, as it is the only independent model.

The second stage involves re-adjusting the probabilistic range that represents the Grey Zone, based on the base model's output probability distribution, followed by updating the Grey Zone models. The adaptive framework takes into account another aspect of the Grey Zone strategy to be changed dynamically, based on the underlying data distribution, which is updating underperforming Grey Zone models. To judge whether a model needs to be updated, the adaptive mechanism relies on the models' performance using a cross-validation method on the updated dataset. Detailed information on the adaptive algorithms is presented in the rest of this section.

7.3.2 AGZEM Framework Construction

This section details the adaptive mechanism developed to update the multi-course early warning framework for at-risk students by performing dynamic changes in the predictive instrument's structure, parameters and configuration. Such an adaptive strategy guarantees the dynamic adaption of new observations when changes occur in the prediction setting. It is important to select a combination of approaches that suits the prediction context. Therefore, we propose an Ensemble Model Adaptive algorithm (EMA) to handle the base and Grey Zone models' updates. An EMA algorithm relies on both structural and combinational adaptive mechanisms to update underlying models. The other proposed algorithm is the Grey Zone Bounds Adjustment Algorithm (GZBA). A GZBA takes care of the changing Grey Zone upper and lower boundaries adaptively. The rest of this section describes both algorithms in detail.

Description of the EMA algorithm

The EMA algorithm is designed to comply with the ensemble model strategy presented in Chapter 6, where the utilised design of the ensemble consists of a fixed-size set of members. Moreover, the ensemble model is associated with each member model, and with a local set of features from the global features space. In other words, each member is an expert in a unique subset of characteristics. The EMA algorithm incorporates structural and combinational adaptive methods to update the predictive ensemble models. Due to restrictions in the ensemble model design where it has a fixed size set of members, the EMA algorithm can perform limited forms of structural adaptive methods, particularly when replacing one existing weak member model with a newly constructed quality model. However, other structural changes may be applicable in conjunction with a combinational updating method, where poorly performing members can be deactivated if associated with a null weighting parameter. Therefore, the EMA algorithm considers updating members' weighting parameters with respect to each member's predictive performance.

The EMA algorithm steps are detailed in Algorithm 1. Its five inputs are: the training and evaluation datasets, a set containing features belonging to each local area, a set containing the initial member models and a set involving all possible weighting combinations. There are three major steps performed in the EMA algorithm, as follows.

- **Step 1:** *Replacing member models step*, for each local area $c_i \in C$, a new model f is trained using the input training dataset (i.e. dataset contains both the past and newly-obtained instances), where the utilised features are selected using a feature selection method from a predefined set of local features. The algorithm compares new and currently existing members using performance evaluation metrics to evaluate each model's quality. The evaluation process is performed using evaluating instances, which have been provided to the algorithm as an input. In the case that a newly-constructed model outperforms the existing one, then it replaces the existing model, otherwise the member model remains with no update. The same procedure is repeated for each one of the eight member models contained in the ensemble learner.

- **Step 2:** *Members weighting step*, where the algorithm optimises the best-fit combination of the weighting parameters dynamically. Members' weights are optimised based on a performance evaluation metric of the combined members. The weighting step aims to weight members based on their performance, wherein each weak member is penalised by receiving a low weight and higher weights are assigned to quality members. This step finishes by updating the ensemble model weighting parameters.

- **Step 3:** *Updating step*, this step is responsible for updating the underlying model after completing the adaptive process. The updated model replaces existing one.

Description of GZBA algorithm

The GZBA algorithm aims to identify optimal cut-off values for the upper and lower boundaries of the Grey Zone with respect to the base model global threshold. The algorithm estimates the upper and lower cut-off values (the optimal Grey Zone upper and lower thresholds) based on a Receiver Operating Characteristics (ROC) graph (Fawcett 2006). ROC works by drawing many points on the graph space starting from the lower left point (0,0) to the top right point (1,1). For starting and finishing points, the predictive model predicts instances to a single class unconditionally, while the upper left point (0, 1) characterises the finest classifications. Therefore, measuring the distances between each point in the ROC space and the top left point indicates the best cut-off point where the shortest distance is the best cut-off point.

The GZBA algorithm identifies the upper and lower Grey Zone boundaries by detecting optimal cut-off values with respect to the area above and below the global threshold, where each area represents an individual class in the prediction space. Analysing each area separately allows us to identify the local probabilistic range, where the prediction classes overlap. Therefore, the GZBA algorithm divides instances into upper and lower clusters, based on the base model's outcomes. Separation criteria for the

Algorithm 1 The EMA algorithm

Input:

$T = \{(x_1, y_1), \ldots, (x_i, y_i), i = 1, \ldots, I\}$, where $y_i \in \{0, 1\}$: training features and target data in batches

$E = \{(x_1, y_1), \ldots, (x_j, y_j), j = 1, \ldots, J\}$, where $y_j \in \{0, 1\}$: evaluation features and target data in batches

$C = \{c_1, \ldots, c_m), m = 1, \ldots, M\}$: set of local features belonging to each category

$G = \{g_1, \ldots, g_m), m = 1, \ldots, M\}$: set of initial model members

$W = \{w_1, \ldots, w_m), k = 1, \ldots, K\}$: set of possible weighting combinations

Step 1: Replacing member models

/* Building models using provided datasets with the help of the feature selection method, then determining whether an existing model should be updated or not */

for each features local category $c_i \in C$ **do**
    $f \leftarrow$ create new model, where features $\in c_i$
    $b\_performance \leftarrow evaluate(g_i, E)$ //compute performance of initial model g_i using evaluation data
    $f\_performance \leftarrow evaluate(f, E)$ //compute performance of new model f using evaluation data
    if $f\_performance > b\_performance$ **then**
        replace initial member by f
    else
        initial model does not need to be replaced
    end if
end for

Step2: Weighting members

/* Optimising best-fit members weights */

for each weighting combination $w_k \in W$ **do**
    compute ensemble model performance based on w_k using evaluation data batches
end for
$E \leftarrow$ *the best-fit members weight combination*

Step 3: Updating

/*Updated ensemble model replaces existing ensemble learner*/

Algorithm 2 The GZBA algorithm

Input:

D, the set of evaluation instances; $f(i)$ is instance prediction probability and $y(i)$ is instance actual label where $y(i) \in \{0, 1\}$

θ, the global threshold;

P_{Lower} and N_{Lower}, the number of positive and negative instances below the global threshold;

P_{Upper} and N_{Upper}, the number of positive and negative instances below the global threshold

Ensure:

$P_{Lower} > 0$, $N_{Lower} > 0$, $P_{Upper} > 0$ and $N_{Upper} > 0$

Step 1: Initialisation

$TP_{Lower} \leftarrow TP_{Upper} \leftarrow 0$
$TN_{Lower} \leftarrow N_{Lower}$
$TN_{Upper} \leftarrow N_{Upper}$

Step 2: Grouping count of positive and negative instances by probability score value

$D_{sotrted} \leftarrow D$ sorted in descending order
/* Q is a queue of objects where each one is a structure of three members, positives and negatives counts and probability score */
$Q \leftarrow \langle\rangle$
positives $\leftarrow$ negatives $\leftarrow 0$
$f_{previous} \leftarrow -\infty$
$i \leftarrow 1$

while $i \leq |D_{sotrted}|$ **do**
    if $f(i) \neq f_{previous}$ **then**
        enqueue (positives, negatives, $f_{previous}$) onto Q
        $positives \leftarrow negatives \leftarrow 0$
    end if
    $f_{previous} \leftarrow f(i)$
    if $y(i)$ is positive instance **then**
        $positives \leftarrow positives + 1$
    else // instance is negative
        $negatives \leftarrow negatives + 1$
    end if
    $i \leftarrow i + 1$
end while

Algorithm 2 The GZBA algorithm (Continued)

Step 3: Compute the distance from each ROC point to the top-left coroner and select the point with the minimum distance

$\theta_{Lower} \leftarrow \theta_{Upper} \leftarrow \theta$
$dist_{Lower} \leftarrow dist_{Upper} \leftarrow 1$

for each object in Q, q **do**
  if $q.score \geq \theta$ **then** /* processing the upper part of the probability distribution */
    $TP_{Upper} \leftarrow TP_{Upper} + q.positives$
    $TN_{Upper} \leftarrow TN_{Upper} - q.negatives$
    $distance = \sqrt{(1 - \frac{TP_{Upper}}{P_{Upper}})^2 + (1 - \frac{TN_{Upper}}{N_{Upper}})^2}$
    if $dist_{Upper} > distance$ **then**
      $\theta_{Upper} \leftarrow q.score$
      $dist_{Upper} \leftarrow distance$
    end if
  else /* processing the lower part of the probability distribution where the probabilities are below the threshold */
    $TP_{Lower} \leftarrow TP_{Lower} + q.positives$
    $TN_{Lower} \leftarrow TN_{Lower} - q.negatives$
    $distance = \sqrt{(1 - \frac{TP_{Lower}}{P_{Lower}})^2 + (1 - \frac{TN_{Lower}}{N_{Lower}})^2}$
    if $dist_{Lower} > distance$ **then**
      $\theta_{Lower} \leftarrow q.score$
      $dist_{Lower} \leftarrow distance$
    end if
  end if
end for

Outputs: a set contains θ_{Lower} and θ_{Upper}

base model outcomes rely solely on the base model's global threshold, where evaluated instances with computed probabilities of or above the global threshold are assigned to the upper group and the rest are allocated to the lower probability group. Then, the algorithm measures the distance between each possible threshold point and the top left corner to select the optimal local cut-off for each group, where the cut-off point with the shortest distance is nominated. The distance is calculated based on the True Positive (TP) and the True Negative (TN) rates at each ROC point on the graph. The GZBA algorithm utilises a set of inputs as follows: evaluating instances prediction probabilities produced by the underlying base model and its global threshold and, the total number of positive and negative instances located in the upper and lower regions of the probability distribution.

The GZBA algorithm (Algorithm 2) involves three major steps which are:

- **Step 1:** *Initialisation step*, the algorithm begins by initialising the parameters where the upper and lower True Positive value (TP) is set to 0. Furthermore, the upper and lower True Negative value (TN) is set to the total number of negative instances corresponding to each cluster.

- **Step 2:** *Grouping instances by probability score value step*, it is unusual for multiple instances to have the exact same outputs as human behaviours are not identical, but such a rare event may occur. Therefore, to overcome this problem, the algorithm groups instances with identical prediction scores in one structure that holds three members: the portability score, total actual positive instances and total number of actual negative instances with the computed prediction probability that matches the object score. The algorithm takes advantage of a descending order to make the grouping process more efficient. A descending ordering of prediction scores makes instances with equal scores adjacent. The algorithm keeps tracing probability scores by comparing previous and current scores to determine when to stop counting instances and enqueue a created object into a queue Q.

- **Step 3:** *Evaluate and select the ROC point step*, in the final step, the algorithm computes the distance between each point in the ROC space and the top-left point of the ROC graph where the point with the shortest distance represents the best cut-off point for the given set of instances. Taking the descending order of the previous step queue Q into the account, any positive instance assigned to the positive class with respect to a specified threshold remains predicted as a positive for any lower threshold. In the same way, any negative instance predicted as a negative with respect to a specified threshold remains predicted as a negative for any higher threshold. Consequently, the algorithm is required to process one object at the time and keep updating TP and TN for every iteration where TP changes detrimentally and TN changes incrementally. Then, the distance to the upper left point is calculated with respect to sensitivity and specificity metrics at each threshold point using the following formula:

$$distance = \sqrt{((1 - sensitivity)^2 + (1 - specificity)^2)} \quad (7.1)$$

After measuring the distance for the top-left corner to each the ROC point, the algorithm preserves the minimum distance and corresponding cut-off value. This step is performed for the upper and lower range of probabilities individually, only using instances that fall in the underlying range. Finally, the algorithm returns a pair of values that contain the identified optimal upper and lower cut-offs that represent the Grey Zone boundaries.

7.4 Experimental Study

To test the feasibility of the proposed adaptive strategy in forecasting at-risk students setting, we evaluated the effects of the developed adaptive framework on the updated predictive models' performances against the performance baseline model. The underlying predictive model's design matches the exact same design used to implement the multi-course early warning framework developed in Chapter 6. The baseline model is the initial model which remains static, with no modification throughout the experimental study. Furthermore, experiments were carried out over two adaptive scenarios, with and without performing the forgetting mechanism that removes irrelevant instances from the historical dataset when new adaption data batches are fed to the adaptive framework to trigger the adoption process.

Datasets used during the adoption process belong to twelve courses, where data was drawn from each course and held in an independent data batch. The remaining data, which belonged to a single course, were used to build an initial baseline predictive model, which represents the starting model. The adaptive process was performed incrementally, using adaption dataset sizes where a 10-fold cross-validation method was used to split the underlying adaptation data into training and evaluation datasets. An AUC metric is used to evaluate and compare the changes in the updated models' performances, alongside the baseline model's performance for each study week of the 12-week long semester. The adaptive framework was implemented using Java, with the help of a Weka machine learning library that was used to construct logistic regression models.

7.4.1 The Dataset

The data was collected from thirteen blended computer science courses taught at the University of Adelaide, Australia, over the first and second semesters between 2012 and 2016. Each semester consists of 12 core study weeks interspersed with a 2 week mid-semester break and optional teaching weeks. Out of the thirteen collected courses, eight courses were only made available for undergraduate students, while the rest were offered to students at both undergraduate and postgraduate levels. Students' interactions and participation data was drawn from the Moodle VLE used in the School of Computer Science.

7.4.2 Experimental Setup

Adaptive strategies were implemented in Java following the proposed AGZEM framework, which consists of two major components: data preparation and predictive model update components. The latter component was responsible for updating the predictive base and Grey Zone models alongside re-configuring the Grey Zone parameters.

The data preparation component received the newly-obtained dataset and handled all the data-related processes to be properly organised and pre-processed for analysis. The process began by validating and cleaning the obtained records in the dataset automatically. The process omits all records performed outside the semester's official time-frame, alongside those records belonging to irrelevant VLE users. Furthermore, the posts' contents were cleared of unwanted, non-ASCII characters and noise texts, while the meanings expressed in the posts are maintained and were not affected by the cleaning process.

Moreover, the data preparation component performs time-series generation to group students' actions in weekly data blocks. The component also performs the students' features extraction process from the prepared data. Then, the component handles any outlying observations and transforms the data points. Data transformation is performed by applying min-max normalisation and logarithmic transformation methods. Moreover, the data component is responsible for integrating the newly-prepared adaption dataset with the archived dataset, which is retrieved from a special data warehouse.

The other component is the model update that handles the process of adapting data batches to ensemble learner members using the EMA algorithm described in Section 7.3.2, as well as reconfiguring the utilised Grey Zone parameters using the GZBA algorithm described also in Section 7.3.2. The EMA algorithm is designed to perform the adaptive process on a fixed-size ensemble learner in which each member is an expert in a local area of the features space. The adaptive ensemble process deploys on three stages: updating the base model, re-identifying the Grey Zone boundaries and updating the corresponding weekly and prior-to-break Grey Zone models.

In the predictive models' update component, the EMA algorithm is responsible for updating the underlying predictive models. At the beginning, the EMA algorithm updates the base ensemble model member structurally, using a heuristic replacement strategy where a 10-fold repeated random cross-validation method is utilised. Instances are randomly divided into one of the 10 subsets of data. The process is repeated 10 times where the kth fold is used for validation and the rest of the folds are used to train member models.

In every iteration, a wrapper subset features selection method (Kohavi & John 1997) is performed to choose the most important subsets of features from pre-identified local sets of features associated with each member model. The newly-developed logistic regression member replaces the initial existing member when it performs better, where both members are evaluated on the same data fold. Following the structural adaptive stage of the eight members, the EMA algorithm optimises the best-fit combination members' weighting parameters. It compares the quality of the best performing newly-developed base model and initial base models by averaging the AUC measurement results across the prediction weeks. In cases where the newly-developed base model outperforms the existing ones, it replaces them, otherwise the initial models remain in use.

The next adaptive algorithm is concerned with identifying Grey Zone boundaries based on the updated base model. The GZBA algorithm is utilised to fulfil this task with the help of a total adaption dataset, where it detects the optimal upper and lower boundaries of the Grey Zone dynamically. Following this adaptive stage, the Grey Zone models are updated using the same procedure utilised to update the base model. However, instead of using a 10-fold cross-validation method to split the dataset, only instances that fall in the identified Grey Zone are assigned to the validation dataset, while the remaining instances are assigned to the training dataset.

Moreover, other data-related challenges arose in the experimental study. Firstly, when the adoption process is performed with a forgetting mechanism, the adaption data are tested for statistically significant changes in data distribution using a t-test where the significant parameter is 0.05. Once a null hypothesis is rejected, the forgetting mechanism is performed by removing instances that are not relevant to the new concept. On the other hand, instance removal may cause an imbalance in the class distribution alongside any unbalanced class involved in the adaption dataset. This situation occurs when the underlying dataset has a significantly different number of instances belonging to each target class. This event may negatively affect the model's overall performance. Popular approaches to overcome this problem in machine learning setting are the under-sampling instances belonging to the majority class and over-sampling of the minorities. In this experimental study, in the event of having an imbalanced class distribution in the training folds, the class distribution is balanced using the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al. 2002). SMOTE is a well-known approach to tackle the imbalanced dataset problem by creating new synthetic instances between the closest neighbours from minority classes.

7.4.3 Results

A set of experiments was carried out to evaluate the usefulness of the proposed adaptive algorithmic framework using a sequence of adaptation data batches. The adaptive process involves updating the predictive model's structure and combinational parameters mechanisms, alongside updating the Grey Zone configuration dynamically. Adaption data batches were sorted in chronological order, so that data batch belonging to earlier-offered courses are utilised first. Moreover, we assume that the students' final performances become available at the end of each semester. Therefore, each adaptation course contains students' semester-long digital learning traces.

We evaluated the impact of the adoption process on an initial predictive model's performance over two adaptive scenarios using 12 adaptation courses: (scenario A) appending the adaption batches to the full size of the historical training dataset and (scenario B) utilising a forgetting mechanism to remove irrelevant instances from the past training dataset. In the latter scenario, the forgetting mechanism executes whenever a new adaption dataset has statistically significant differences in data distribution when compared with the existing training dataset.

The overall changes in the trends in the initial and updated models' qualities in terms of an averaged AUC metric over the 12 adaption datasets are illustrated in Figure 7.2. It can be seen that after every deployment of an adaptive mechanism, the updated predictive models change their properties to cope with the changes in the prediction concepts and improve their performance, when compared with the baseline model's performance. Generally, the developed adaptive strategy affected the predictive model's prediction quality noticeably when compared with the baseline model's predictions. It was found that adaptive scenario (B) results in significant enhancements in the models' performances where irrelevant instances belonging to the historical dataset were forgotten. Irrelevant samples are those which have significantly different concepts from the most recent concept. Moreover, executing an adaptive approach over adaptive scenario (A) results in decent improvements in the overall updated models' quality. However, over other adaptive scenarios, when a sudden drift occurs in the adaption concept, the adaptive mechanism fails to maintain or enhance the updated predictive models' performance.

Table 7.1 shows the evaluation results of different sequences of adaption datasets over both adaptive scenarios, alongside the baseline models' performances. The results of deploying adaptive learning approaches are computed by averaging the weekly predictions over the mean of the weekly cross-validation folds in terms of the AUC metric. In terms of the first adaptive scenario (A), where we deploy an adaptive mechanism with no forgetting mechanism, the model's response to changes in data distribution enhances its overall quality following each adaptation data batch. The updated model's performances outperform the baseline model performances by 7 percent on average following each adaptive process. The updated model's improvements illustrate the influence of the adaptive strategy on the predictive models to accommodate new concepts involved in the adoption data batches. However, although adaptive

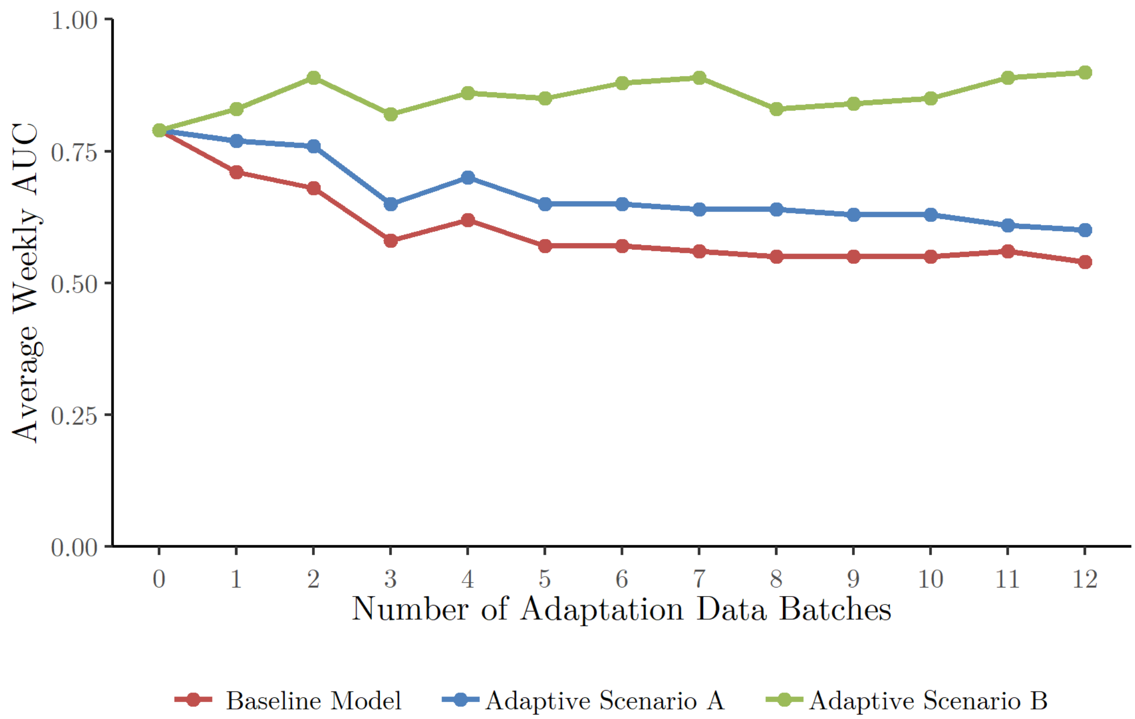


Figure 7.2: An illustration of the changing trends in the predictive models' qualities in terms of the averaged AUC.

	Baseline Model	Scenario A	Scenario B
	Average AUC	**Average AUC**	**Average AUC**
Initial Model	0.79	0.79	0.79
Adaptation course 1	0.71	0.77	0.83
Adaptation course 2	0.68	0.76	0.89
Adaptation course 3	0.58	0.65	0.82
Adaptation course 4	0.62	0.7	0.86
Adaptation course 5	0.57	0.65	0.85
Adaptation course 6	0.57	0.65	0.88
Adaptation course 7	0.56	0.64	0.89
Adaptation course 8	0.55	0.64	0.83
Adaptation course 9	0.55	0.63	0.84
Adaptation course 10	0.55	0.63	0.85
Adaptation course 11	0.56	0.61	0.89
Adaptation course 12	0.54	0.6	0.9

Table 7.1: Evaluation of the results in terms of the AUC accuracy results on incremental adaptation batches where the results are computed by averaging the weekly predictions over the mean of the weekly cross-validation folds.

scenario (A) was able to update the models in a way that maintains producing higher levels of prediction quality, the overall models' performance tends to fall throughout the experiments after receiving new data batches in most cases. The falling pattern in the updating models' performances can be explained by the stronger effects of old concepts on the updated models' learning patterns compared with the influence of new concepts, since the majority of training instances belong to the historical dataset drawn from the older courses.

Furthermore, the developed adaptive mechanism utilises a variety of features drawn from different online learning aspects. In some cases, these features are not sufficient to accommodate all the drifts in the prediction concept due to the major differences in learning characteristics' distributions of the individual courses, which can lead to variations in performance when using an adaptive approach. For example, the significant drift in adaption courses' 3 characteristics lead to a noticeable variance in adaption quality over both adaptive scenarios.

In the second adaptive scenario, scenario (B), we address the problems related to aggregating multiple different concepts in the dataset by removing data points irrelevant to the new concept from the existing dataset. Consequently, the adaptive mechanisms give a much better performance in terms of coping with the changes in data distribution, even when the concept is modified significantly. Processing the adaptive strategy with a forgetting mechanism allows the updated models to maintain a significantly higher performance against the baseline model performance, by up to 36 percent. The updated models have performance improvements after executing most of the adaption batches throughout the experimental study. Over scenario (B), after feeding the adaptive mechanism with 12 data batches, the updated models reached an averaged weekly accuracy level of 0.90 AUC points, where the initial model's performance was 0.79 AUC points, which shows the effect of the forgetting mechanism on the models' learning patterns. On the other hand, in limited cases, the underlying adaptive strategy fails to improve or even maintain similar prediction performance levels in the updated models, counter to prior adoption models.

Adoption courses from 1 to 6 were collected from courses offered in 2012 in semesters 1 and 2, however, after the third adaption batches, the predictive models' performances dropped over both adaptive scenarios, alongside the baseline model's performance. The models' performance decreases were motivated by the significant drift in the learning concepts contained in the recently-fed data batch. The next 5 adaption courses were taught in 2013, where baseline model and updated models with no forgetting mechanisms result in steady prediction rates, while the other adaptive scenario produces overall better quality updated models. The last adaption batch was drawn from a course taught in 2016. Utilising the last adaption batch results in a slight drop in the performance of the baseline model, alongside the model updated using the adaptive scenario A strategy, while resulting in improving model performance when the model is updated with a forgetting mechanism.

Finally, the experimental evaluation results show the advantage of using the proposed adaptive algorithmic framework to cope with changes in data distribution, alongside improving the predictive models' performances. The adaptive mechanism allows the predictive models to achieve significantly better predictions compared with the static baseline model's performance. Furthermore, integrating the data forgetting mechanism results in an overall faster adaptation of new concepts involved in the recently-obtained data distribution, as well as significant improvement rates comparing to adaptive scenario (A).

7.5 Summary

Assessments of computerised predictions of students' academic performance have recently grown rapidly in the literature. Much of the earlier work constructs predictive models in static development environments, which makes fixes the underlying predictive instruments, leaving them without the ability to handle any changes which may occur in prediction setting. This fact raises concerns about the model's scalability and ability to cope with future changes in learning behaviours over time. Therefore, there is a gap in developing and employing adaptive mechanisms to enhance predictive instruments' abilities to cope dynamically with changes in prediction concepts.

Existing work in the research field relies on static environments to develop student performance predictive instruments, where models tend to be trained on historical information and remain fixed with no updates. This fact raises issues regarding the capacity of these instruments to cope with any changes that may occur in the prediction concept. However, utilising adaptive mechanisms can lead to improvements in the quality of the subject predictive instruments' outcomes, alongside the ability to adapt to drifts that may occur in the prediction concepts due to changes in students' learning patterns, as a result of changes in the prediction environment or modifications in their learning behaviours.

In this chapter, we introduced an adaptive framework to handle the adaptive process dynamically which is applicable to the multi-course early warning framework. The proposed AGZEM algorithmic framework implements EMA and GZBA algorithms to update the underlying models' dynamically by updating the models using recently-obtained datasets. The EMA algorithm integrates multiple adaptive approaches on different ensemble modelling levels, where it replaces poorly performing experts at model member level and optimises members' weighting parameters on the outputs' combinational level. The other proposed algorithm, the GZBA algorithm, re-identifies Grey Zone boundaries dynamically, based on the changes arising in the outputs' probabilities distribution, as computed by the updated base model.

Moreover, we conducted experimental studies to examine the impact of adaptive processes on fresh testing datasets. The experimental results reveal the usefulness of the proposed adaptive strategy in allowing the underlying predictive instruments to cope with changes in prediction concepts practically, when associated with a forgetting mechanism which leads to faster and better adaptations.

Chapter 8

Conclusion and Future Direction

In this thesis we focused on the early detection of students who are potentially at-risk of failing or dropping out of academic courses in a higher education blended learning setting. The identification of such students relies on students' interactions with the online learning components offered on the course's VLE module and participation in course discussion forums data. A range of quantitative, qualitative and social analysis approaches were performed on the collected datasets drawn from thirteen blended learning courses offered by the School of Computer Science (N = 1,476 enrolments). Moreover, we proposed and evaluated novel Grey Zone modelling to enhance the efficiency and reliability of the binary predictive instruments.

Furthermore, the extracted online learning characteristics are utilised to develop an exemplar multi-course early warning framework for at-risk students, where the underlying predictive instrument follows the proposed Grey Zone design. The developed predictive multi-course framework was evaluated using unseen evaluation datasets. Additionally, we developed an adaptive framework and algorithms applicable to the proposed Grey Zone design, which allow the predictive model to cope with any drifts that may occur in the prediction concept due to changes in students' learning patterns, as a result of changes in the prediction environment or modifications in their online learning behaviours.

This chapter summarises the work included in this thesis, alongside presenting the key contributions involved in each section of the thesis in Section 8.1. Moreover, Sections 8.2 and 8.3 discuss the primary research questions and limitations of this study, respectively while Section 8.4 suggests future research directions which are d to deliver greater focus on developing and utilising adaptive mechanisms in educational research contexts, alongside developing and evaluating personalised intervention strategies for students who are at risk of failure or drop-out.

8.1 Summary

The first chapter provides a brief introduction of the problem of identifying students who might be at academic risk. In addition, it discusses the motivation for and applications of detecting underperforming students in terms of enhancing individual students' performances and the positive effects of identifying such students on overall higher education institutions' outcomes. Furthermore, we discuss the challenges stemming from relying solely on online learning activities, in blended learning, hybrid off-line and online courses.

Chapter 2 provides a comprehensive review of the literature that investigates past efforts regarding predicting student retention and academic performance in higher education contexts. We began this chapter by describing early efforts to identify students' characteristics and correlate them with retention rates and discussed how modern technologies can offer a new source of valuable data that reflects students' learning progress. Digital learning traces allow us to build more reliable and accurate prediction instruments and achieve higher quality and greater accuracy in our prediction results.

The first portion of Chapter 2 presents different data sources and analysis approaches utilised by education researchers to extract predictive variables of students' performance. Furthermore, we cover various attempts to use a range of prediction methods that employ students' personal characteristics and learning performance factors to develop predictive instruments targeting a variety of forms of academic performance, including student attrition, academic risk status, course final marks and assessment grades. Moreover, this chapter presents various successful initiatives to develop and utilise early warning instruments at an institutional level, alongside the positive impact of these instruments on overall academic outcomes. Additionally, we summarise several aspects of the existing works in the literature on blended and online learning settings, including prediction types, prediction methods, sizes and sources of utilised populations, alongside development and evaluation methodologies.

Chapter 3 presents a summary of the gap identified in the literature, alongside describing efforts in this thesis to bridge the identified gap and answer research questions. The effort performed in this work can be divided into four main categories: proposing and evaluating new features extraction methods, proposing a novel technique to enhance the performance of binary predictive instruments, which are called Grey Zone design, developing and evaluating an exemplar Multi-Course Early Warning Framework to detect at-risk students and developing an adaptive solution to allow the underlying predictive instrument to cope with any changes may arise in predicting concepts dynamically.

Chapter 4 aims to introduce the reader to the collected data, data preparation methods, and the CoreNLP toolkit used to fulfil language-based tasks, a logistic regression prediction approach and the evaluation metrics which are utilised throughout this thesis. Moreover, in this study we proposed an automated approach that weighs identified sentiments based on accompanying adverbs' strengths. Due to the absence of a mechanism to indicate the strength of English adverbs, we built a digital adverbial strength dictionary, which is one of the main contributions of this work. The development process for the adverbs strength dictionary is also described in detail in Chapter 4.3. Furthermore, this chapter describes a collection of 53 variables which are extracted using a range of analysis approaches which are used as predictors of students' academic risk status, alongside tools and approaches used to extract predictions and build, update and evaluate predictive instruments.

Chapter 5 presents two key contributions in addition to introducing an automated process for extracting and weighting students' sentiments as expressed in their posts on course discussion boards. The first key contribution, presented in Chapter 5, covers evaluating multiple predictors extracted from discussion forum participation data. Predictors were extracted based on various analysis techniques, including weighted sentiment strength approaches, SNA and trending analysis, to measure the degrees of fluctuation in each participant's characteristics over the study weeks. We employed various well-known machine learning features selection methods to rank the features' predictive power and determine the most influential predictors. The results show that predictors derived from weighted sentiment approaches of student-generated textual-based contents present the majority of the top-ranked discussion forum predictors, where a semester-appended sentiment strength feature is ranked as the top predictor.

Moreover, the second core contribution presented in Chapter 5 is related to the proposed novel Grey Zone strategy, used to enhance the performance of the binary predictive instruments. The proposed strategy works by identifying a probabilistic range where the underlying base model fails to provide quality predictions and utilises an alternative Grey Zone predictive model to predict instances falling in the Grey Zone. A set experiments was carried out to examine the usefulness of Grey Zone modelling. In these experiments we developed early predictive models of at-risk students based on discussion forum data analysed using traditional and Grey Zone predictive strategies. The experimental study shows that using the Grey Zone approach results in a noticeable overall improvement of the predictive models' performance over the traditional prediction strategy by improving the overall weekly model performance by up to 13 percent in terms of the AUC metric and by up to 25 percent in terms of overall accuracy measures.

However, the early predictive models developed in Chapter 5 are limited to predicting outcomes for those students who participate in course discussion forums. Given the fact that participating in such a communication tool is typically voluntarily, in Chapter 6, we extend our work by accommodating predictors extracted from other online learning activities to generalise the application of the predictive models and improve the quality of their predictions.

Chapter 6 presents the development and evaluation of an exemplar Multi-Course Early Warning Framework of at-risk students, which is also a key contribution of this thesis. The framework combines predictors extracted from online learning activities and online discussion forum participation data to detect at-risk students with the help of a Grey Zone strategy. This chapter presents the specifications for a fixed-size ensemble predictive modelling design used to build underlying predictive instruments, where each model's member is an expert in their local area of the global features space.

The other main contribution presented in Chapter 6 is to evaluate the developed multi-course early warning framework's performance using an unseen dataset and locating the optimal intervention timing. Employing a fresh evaluation dataset, we examined the predictive framework's performance when it predicts future events where the evaluation dataset is drawn from four heterogeneous courses (N = 319 enrolments) in terms of their distribution of online activities. The evaluation results revealed that the framework was able to achieve over 0.92 AUC points across most of the evaluation courses and over 87 percent in terms of the overall prediction accuracy. In addition, the careful analysis of the weekly prediction quality indicates week 3 as being the optimal week to establish the provision of additional, targeted support for at-risk students. The results show that, at week 3, the predictive framework was able to achieve quality predictions over 0.8 AUC points in the aggregated testing dataset.

Chapter 7 proposes an adaptive framework applicable to the Grey Zone design to allow the predictive model to cope with any changes that may occur in the prediction space over time or due to changes in the prediction settings. This chapter reviews popular adaptive mechanisms used in ensemble modelling contexts and proposes other adaptive approaches to change the properties of the Grey Zone design dynamically. Moreover, it presents a detailed description of the Adaptive Grey Zone Ensemble Model (AGZEM) framework developed as part of this work. Furthermore, it describes the Ensemble Model Adaptive (EMA) algorithm and the Grey Zone Boundaries Adjustment (GZBA) algorithm involved in the AGZEM framework. The experimental study was analysed in this chapter to evaluate the usefulness of the proposed adaptive framework and algorithms. The adaptive process was deployed over two adaptive scenarios. The adaptive scenarios involve deploying adaptive mechanisms with and without utilising a forgetting instrument for historical data instances, where the forgetting mechanism is utilised whenever a statistically significant change is detected in the data distributions. The results illustrate the practicality of the proposed adaptive framework, and it capacity to allow the underlying updated predictive models to cope with changes in the prediction concepts. Additionally, the results show that integrating the forgetting mechanism for irrelevant historical data leads to faster and better adoption outcomes.

8.2 Thesis Research Questions

The primary research questions of this thesis were:

RQ 1: What are the most influential student online discussion forum participation predictors for students who are at-risk in a blended learning setting?

In Chapter 5, various online discussion forum participation predictors were examined where the objective was identifying the most influential discussion forum predictors of students who were at-risk of not completing their academic courses successfully. Underlying academic risk predictors cover a variety of learning and social aspects gained from investigating student-generated textual content, social characteristics and participation patterns within online discussion forums. Table 5.4 shows the top ve most influential predictors using ve well-known machine learning feature selection approaches. The overall ranking result reveal that the language-based predictors are the most significant discussion forum predictor, followed by social aspects.

RQ 2: What technology is needed to enhance the ability of the predictive model to produce reliable predictions of students who are at-risk?

Relying on the traditional decision-making methodology is common practice to interpret predictive instruments outcomes in the context of identifying academic performance. However, Chapter 5 introduces the novel Grey Zone decision-making design to improve the quality of the binary classifiers. The proposed design suggests further investigation for students for whom their calculated probability falls within pre-defined boundaries. Initial comparison of the experimental results shows promising improvements in predictive instruments performance where applying the Grey Zone design over the traditional decision-making strategy improves the overall weekly predictive instruments accuracy by up to 25 percent. Furthermore, experimenting with the proposed Grey Zone design resulted in providing enhanced overall predictive instruments classification quality by providing higher recall and precision on average.

However, at very limited occasions, the performance of the prediction instruments were dropped in terms of prediction accuracy while achieving better predictions qualities in terms of distinguishing actual instance classes measured by the AUC metric.

RQ 3: How can a reliable early warning framework of at-risk students that supports multiple courses be developed using VLE interactions and discussion forum data in a blended learning setting?

Developing a multi-course early warning framework of at-risk students is the subject of Chapter 6. In this chapter, we developed an exemplar predictive framework that detects students who might be at failure or attrition risk, in a weekly manner powered solely by VLE interactions and discussion forum data. The framework was built using an ensemble modelling strategy, where the ensemble predictive model consists of eight weighted members; each individual member is developed using a unique set of features belonging to a single category. Each features category combines a unique subset of features from the global space. Furthermore, the underlying framework was implemented with the help of a Grey Zone decision-making strategy to improve the prediction accuracy.

Evaluation results of the developed framework are presented in Section 6.6 where the framework is evaluated with four entire unseen blended learning courses datasets (N = 319 enrolments). Experimental results illustrate the predictive framework's ability to predict future events with high classification rates across the majority of the evaluation courses where the framework's top performance ranged from 0.81 to 0.94 AUC points across the testing courses. In terms of accuracy metrics, the framework obtained its best performance, between 77 and 90 percent, across individual courses in the evaluation dataset. Finally, the developed framework was able to provide reliable predictions as early as week 3 of the semester.

RQ 4: What are the adaptive strategies that can be used to allow the proposed framework to cope with any changes that may occur in the prediction space dynamically to maintain its ability to produce reliable predictions?

The multi-course framework described in Chapter 6 was developed under a static development environment which leaves its predictive instruments without the ability to handle changes that may occur in prediction setting. Therefore, in Chapter 7, we introduced adaptive strategies that allow the predictive instrument to cope with any changes that may occur in the prediction space dynamically.

Chapter 7 describes the proposed AGZEM algorithmic framework which implements the EMA and GZBA algorithms to update the underlying predictive instruments dynamically by updating the models using recently-obtained datasets. The EMA algorithm integrates multiple adaptive approaches on different ensemble modelling levels, where it replaces poorly performing experts at model member level and optimises members' weighting parameters on the outputs' combinational level. The other proposed algorithm, the GZBA algorithm, re-identifies Grey Zone boundaries dynamically, based on changes arising in the outputs' probability distribution, as computed by the updated base model.

Experimental results presented in Section 7.4.3 confirm the usefulness of the proposed adaptive strategy in allowing the underlying predictive instruments to cope with changes in prediction concepts practically. When associated with a forgetting mechanism this leads to faster and better adaptations in predicting students' performance setting.

8.3 Limitations of Study

Throughout this thesis, students VLE interactions and participations data are the major vehicle for detecting students who are at academic risk. Although relying only on such data sources resulted in decent prediction outcomes, being limited to one source of data reduces the ability of predictive instruments to employ further academic risk characteristics which may exist in other data sources such static or academic sources.

Furthermore, this study is limited to data collected from courses offered at a single school in one university; therefore the researchers were not able to examine the performance of the adaptive mechanisms under different educational environments.

Finally, this study is limited to utilising the power of logistic regression to preform prediction tasks based on the literature analysis rather verifying its performance or comparing it with other commonly used methods.

8.4 Future Direction

Throughout this thesis, we developed and employed multiple approaches to analyse students' online learning patterns and subsequently predict students who are at risk of failure in blended learning setting. In terms of adaption, we develop adaptive mechanisms that enable the predictive model to adapt to changes in the prediction space. There are several prospects for extending the work demonstrated in this thesis including:

- Extending the dataset: utilising a dataset drawn from multiple different institutions may be a useful step to examine the ability of the proposed adaptive framework and algorithms to cope with changes in prediction concepts due to changes in the prediction environment.

- Altering the proposed adaptive framework: the adaptive framework and algorithms can be altered to involve additional adaptive mechanisms and other supporting methods. Supporting approaches might be useful to enhance the adaptive process by treating data-related problems including imbalanced classes problems, evaluating alternative methods to distinguish changes in data distributions, examining alternative methods to detect and remove irrelevant instances in the adaptation dataset or substituting data pre-processing techniques to optimise the quality of the training data.

- Developing an intervention strategy: it might be useful to provide students in danger of academic failure with personalised, proactive intervention actions or feedback. Analysing students' learning data to identify their weakness and consequently design a personalised interventions plan can be a useful step towards achieving the objective of enhancing each student's learning outcomes. Achieving such a target requires the development of a reliable and efficient automated mechanism to plan for appropriate and personalised support for students at risk, without affecting the lecturers' workload.

- Studying the impact of the multi-course early warning framework, along with a proactive intervention strategy: investigating the effects of detecting potential at-risk students and delivering personal interventions on actual students' performance to analyse the degree of impact of such actions in a real-life setting. Furthermore, collecting the lecturers' feedback about utilising such a predictive framework and correction plan may open new avenues for future research.

Appendix A

Statistics of Collected Data

	Viewing Module		Viewing Resources		Creating Posts		Viewing Posts	
	Total	**Percent**	**Total**	**Percent**	**Total**	**Percent**	**Total**	**Percent**
Course 1	5231	36.1%	3830	26.5%	254	1.8%	5158	35.6%
Course 2	3162	38.4%	571	6.9%	213	2.6%	4287	52.1%
Course 3	14310	31.2%	7589	16.6%	897	2.0%	23054	50.3%
Course 4	2180	43.6%	195	3.9%	171	3.4%	2457	49.1%
Course 5	6856	47.5%	1272	8.8%	211	1.5%	6098	42.2%
Course 6	8325	57.1%	680	4.7%	90	0.6%	5486	37.6%
Course 7	3932	40.5%	1245	12.8%	201	2.1%	4342	44.7%
Course 8	1431	42.0%	770	22.6%	69	2.0%	1137	33.4%
Course 9	3740	33.2%	1789	15.9%	232	2.1%	5507	48.9%
Course 10	4362	32.1%	6775	49.8%	120	0.9%	2337	17.2%
Course 11	6120	43.5%	3187	22.6%	149	1.1%	4627	32.9%
Course 12	6419	31.7%	5230	25.8%	347	1.7%	8258	40.8%
Course 13	10964	52.4%	3466	16.6%	256	1.2%	6237	29.8%

Table A.1: Students' virtual learning activities' distribution across the collected courses.

	Viewing Module		Viewing Resources		Creating Posts		Viewing Posts	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Course 1	373.64	166.87	273.57	112.24	18.14	19.57	368.43	441.80
Course 2	225.86	96.07	40.79	21.36	16.38	21.32	306.21	434.65
Course 3	1022.14	263.07	542.07	136.95	64.07	45.75	1646.71	767.59
Course 4	155.71	45.95	13.93	11.76	12.21	13.73	175.50	129.07
Course 5	489.71	218.54	90.86	83.16	15.07	13.01	435.57	416.51
Course 6	594.64	229.43	48.57	72.22	6.43	10.04	391.86	283.29
Course 7	280.86	115.83	88.93	41.35	14.36	27.08	310.14	448.84
Course 8	102.21	67.77	55.00	45.67	4.93	5.53	81.21	103.03
Course 9	267.14	137.31	127.79	51.16	16.57	20.72	393.36	542.41
Course 10	311.57	92.61	483.93	173.58	8.57	8.42	166.93	143.59
Course 11	437.14	246.82	227.64	210.45	10.64	12.12	330.50	408.80
Course 12	458.50	191.14	373.57	196.72	24.79	27.89	589.86	716.91
Course 13	783.14	339.44	247.57	92.82	18.29	16.71	445.50	384.40

Table A.2: Statistical analysis of the weekly virtual learning activities across the collected courses.

	Viewing Module		Viewing Resources		Creating Posts		Viewing Posts	
	Mean	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**
Course 1	2.44	1.09	1.79	0.73	0.12	0.13	2.41	2.89
Course 2	2.69	1.14	0.49	0.25	0.18	0.25	3.65	5.17
Course 3	7.92	2.04	4.20	1.06	0.50	0.35	12.77	5.95
Course 4	1.26	0.37	0.11	0.09	0.10	0.11	1.42	1.04
Course 5	4.71	2.10	0.87	0.80	0.14	0.13	4.19	4.00
Course 6	4.07	1.57	0.33	0.49	0.04	0.07	2.68	1.94
Course 7	1.95	0.80	0.62	0.29	0.10	0.19	2.15	3.12
Course 8	0.77	0.51	0.42	0.35	0.04	0.04	0.62	0.78
Course 9	1.96	1.01	0.94	0.38	0.12	0.15	2.89	3.99
Course 10	3.85	1.14	5.97	2.14	0.11	0.10	2.06	1.77
Course 11	5.68	3.21	2.96	2.73	0.14	0.16	4.29	5.31
Course 12	6.03	2.52	4.92	2.59	0.33	0.37	7.76	9.43
Course 13	9.21	3.99	2.91	1.09	0.22	0.20	5.24	4.52

Table A.3: Statistical analysis of the weekly virtual learning activities per student across the collected courses.

References

Adamopoulos, P. (2013), What makes a great MOOC? an interdisciplinary analysis of student retention in online courses, *in* 'Proceedings of the Thirty Fourth International Conference on Information Systems (ICIS)', pp. 1–21.

Agnihotri, L. & Ott, A. (2014), Building a Student At-Risk Model: An End-to-End Perspective From User to Data Scientist, *in* 'Proceedings of the 7th International Conference on Educational Data Mining', pp. 209–212.

Agudo-Peregrina, A. F., Iglesias-Pradas, S., Conde-Gonzalez, M. A. & Hernandez-Garcia, A. (2014), 'Can we predict success from log data in VLEs? classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning', *Computers in Human Behavior* **31**, pp. 542–550.

Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A. & Goodrich, V. (2014), Engagement vs Performance: Using electronic portfolios to predict first semester engineering student retention, *in* 'Proceedings of the Fourth International Conference on Learning Analytics And Knowledge', LAK '14, ACM, New York, NY, USA, pp. 103–112.

Ahadi, A., Lister, R., Haapala, H. & Vihavainen, A. (2015), Exploring machine learning methods to automatically identify students in need of assistance, *in* 'Proceedings of the eleventh annual International Conference on International Computing Education Research - ICER 15', ACM Press.

Aldowah, H., Al-Samarraie, H. & Fauzy, W. M. (2019), 'Educational data mining and learning analytics for 21st century higher education: A review and synthesis', *Telematics and Informatics* **37**, 13 –49.

Alippi, C., Boracchi, G. & Roveri, M. (2012), Just-in-time ensemble of classifiers, *in* 'Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN)', pp. 1–8.

Almosallam, E. A. & Ouertani, H. C. (2014), Learning Analytics: definitions, applications and related fields, *in* 'Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)', Springer Singapore, Singapore, pp. 721–730.

Armstrong, J. S. (2012), 'Illusions in regression analysis', *International Journal of Forecasting* **28**(3), pp. 689–694.

Arnold, K. E. & Pistilli, M. D. (2012), Course Signals at Purdue: Using learning analytics to increase student success, *in* 'Proceedings of the 2nd International Conference on Learning Analytics and Knowledge', LAK '12, ACM, New York, NY, USA, pp. 267–270.

Ashenafi, M. M., Riccardi, G. & Ronchetti, M. (2015), Predicting students' final exam scores from their course activities, *in* 'Proceedings of the 2015 IEEE Frontiers in Education Conference (FIE)', pp. 1–9.

Ashenafi, M. M., Ronchetti, M. & Riccardi, G. (2016), Predicting student progress from peer-assessment data, *in* 'Proceedings of the 9th International conference on Educational Data Mining', pp. 270–275.

Astin, A. W. (1984), 'Student Involvement: A developmental theory for higher education', *Journal of College Student Development* **25**(4), pp. 297–308.

Azcona, D. & Casey, K. (2015), 'Micro-analytics for student performance prediction leveraging fine-grained learning analytics to predict performance', *International Journal of Computer Science and Software Engineering (IJCSSE)* **4**(8), pp. 218–223.

Bainbridge, J., Melitski, J., Zahradnik, A., Lauria, E., Jayaprakash, S. & Baron, J. (2015), 'Using learning analytics to predict at-risk students in online graduate public affairs and administration education', *Journal of Public Affairs Education* **21**(2), pp. 247–262.

Baker, R. W. & Siryk, B. (1984), 'Measuring adjustment to college', *Journal of Counseling Psychology* **31**(2), pp. 179–189.

Bates, A. W. (2005), *Technology, E-learning and Distance Education*, 2nd edn, Routledge, Abingdon.

Bayer, J., Bydzovska, H., Geryk, J., Obsivac, T. & Popelinsky, L. (2012), Predicting drop-out from social behaviour of students, *in* 'Proceedings of the 5th International Conference on Educational Data Mining', pp. 103–109.

Bin Mat, U., Buniyamin, N., Arsad, P. M. & Kassim, R. (2013), An overview of using academic analytics to predict and improve students' achievement: A proposed proactive intelligent intervention, *in* 'Proceedings of the 5th IEEE Conference on Engineering Education (ICEED)', pp. 126–130.

Binali, H. H., Wu, C. & Potdar, V. (2009), A new significant area: Emotion detection in e-learning using opinion mining techniques, *in* 'Proceedings of the 3rd IEEE International Conference on Digital Ecosystems and Technologies', pp. 259–264.

Bouchachia, A. (2011), 'Incremental learning with multi-level adaptation', *Neurocomputing* **74**(11), pp. 1785–1799.

Boyer, S. & Veeramachaneni, K. (2016), Robust predictive models on MOOCs: Transferring knowledge across courses, *in* 'Proceedings of the 9th International Conference on Educational Data Mining', pp. 298–305.

Buckingham Shum, S. & Ferguson, R. (2012), 'Social learning analytics', *Journal of Educational Technology & Society* **15**(3), pp. 3–26.

Burgos, C., Campanario, M. L., de la Pena, D., Lara, J. A., Lizcano, D. & Martinez, M. A. (2018), 'Data mining for modeling students performance: A tutoring action plan to prevent academic dropout', *Computers & Electrical Engineering* **66**, pp. 541–556.

Bydzovska, H. (2016), A comparative analysis of techniques for predicting student performance, *in* 'Proceedings of the 9th International Conference on Educational Data Mining', pp. 306–311.

Caballe, S., Daradoumis, T., Xhafa, F. & Juan, A. (2011), 'Providing effective feedback, monitoring and evaluation to on-line collaborative learning discussions', *Computers in Human Behavior* **27**(4), pp. 1372–1381.

Cambruzzi, W., Rigo, S. J. & Barbosa, J. L. V. (2015), 'Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach', *Journal of Universal Computer Science* **21**(1), pp. 23–47.

Campbell, J., DeBlois, B. & G. Oblinger, D. (2007), 'Academic analytics: A new tool for a new era', *EDUCAUSE Review* **42**(4), pp. 42–57.

Casey, D. M. (2008), 'A journey to legitimacy: The historical development of distance education through technology', *TechTrends* **52**(2), pp. 45–51.

Cassells, L. (2018), 'The effectiveness of early identification of 'at risk' students in higher education institutions', *Assessment & Evaluation in Higher Education* **43**(4), pp. 515–526.

Cen, L., Ruta, D., Powell, L., Hirsch, B. & Ng, J. (2016), 'Quantitative approach to collaborative learning: performance prediction, individual assessment, and group composition', *International Journal of Computer-Supported Collaborative Learning* **11**(2), pp. 187–225.

Cerezo, R., Sanchez-Santillan, M., Paule-Ruiz, M. P. & Nunez, J. C. (2016), 'Students' LMS interaction patterns and their relationship with achievement: A case study in higher education', *Computers & Education* **96**, pp. 42–54.

Chai, K. E. & Gibson, D. (2015), Predicting the risk of attrition for undergraduate students with time based modelling, *in* 'Proceedings of the 12th International

Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2015)', pp. 109–116.

Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. (2002), 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research* **16**, pp. 321–357.

Cheng, C. K., Pare, D. E., Collimore, L.-M. & Joordens, S. (2011), 'Assessing the effectiveness of a voluntary online discussion forum on improving students course performance', *Computers & Education* **56**(1), pp. 253–261.

Choi, S. P. M., Lam, S. S., Cheong Li, K. & Wong, B. T. M. (2018), 'Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions', *Journal of Educational Technology & Society* **21**(2), pp. 273–290.

Chu, F. & Zaniolo, C. (2004), Fast and light boosting for adaptive mining of data streams, *in* 'Advances in Knowledge Discovery and Data Mining', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 282–292.

Clark, E. E. & Ramsay, W. (1990), 'Problems of retention in tertiary education', *Education Research and Perspectives* **17**(2), pp. 47–59.

Coates, H., James, R. & Baldwin, G. (2005), 'A critical examination of the effects of learning management systems on university teaching and learning', *Tertiary Education and Management* **11**(1), pp. 19–36.

Conijn, R., Snijders, C., Kleingeld, A. & Matzat, U. (2017), 'Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS', *IEEE Transactions on Learning Technologies* **10**(1), pp. 17–29.

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S. & Baker, R. S. (2016), Combining click-stream data with nlp tools to better understand MOOC completion, *in* 'Proceedings of the Sixth International Conference on Learning Analytics & Knowledge', LAK '16, ACM, New York, NY, USA, pp. 6–14.

Dascalu, M., Popescu, E., Becheru, A., Crossley, S. & Trausan-Matu, S. (2016), Predicting academic performance based on students' blog and microblog posts, *in* 'Proceedings of the 11th European Conference on Technology Enhanced Learning (EC-TEL 2016)', Springer International Publishing, Cham, pp. 370–376.

Dawson, S., Jovanovic, J., Gašević, D. & Pardo, A. (2017), From prediction to impact: Evaluation of a learning analytics retention program, *in* 'Proceedings of the Seventh International Learning Analytics & Knowledge Conference', LAK '17, ACM, New York, NY, USA, pp. 474–478.

Department of Education and Training, Australia (2018). viewed 16 November 2018. **URL:** *http://highereducationstatistics.education.gov.au/Default.aspx*

Dietterich, T. G. (2000), Ensemble methods in machine learning, *in* 'Proceedings of the International Workshop on Multiple Classifier Systems (MCS)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–15.

Dodge, B., Whitmer, J. & Frazee, J. P. (2015), Improving undergraduate student achievement in large blended courses through data-driven interventions, *in* 'Proceedings of the Fifth International Conference on Learning Analytics And Knowledge', LAK '15, ACM, New York, NY, USA, pp. 412–413.

Dominguez, M. M., Bernacki, M. L. & Uesbeck, P. M. (2016), Predicting stem achievement with learning management system data: Prediction modeling and a test of an early warning system, *in* 'Proceedings of the 8th International Conference on Educational Data Mining', pp. 589–590.

El-Halees, A. (2011), Mining opinions in user-generated contents to improve course evaluation, *in* 'Proceedings of the 2nd International Conference on Software Engineering and Computer Systems (ICSECS 2011)', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 107–115.

Elwell, R. & Polikar, R. (2011), 'Incremental learning of concept drift in nonstationary environments', *IEEE Transactions on Neural Networks* **22**(10), pp. 1517–1531.

Er, E., Gomez-Sanchez, E., Bote-Lorenzo, M. L., Dimitriadis, Y. & Asensio-Perez, J. I. (2017), Predicting peer-review participation at large scale using an ensemble learning method, *in* 'Proceedings of the Learning Analytics Summer Institute', pp. 55–62.

Everett, J. E. & Robins, J. (1991), 'Tertiary entrance predictors of first-year university performance', *Australian Journal of Education* **35**(1), pp. 24–40.

Falkner, N. J. & Falkner, K. E. (2012), A fast measure for identifying at-risk students in computer science, *in* 'Proceedings of the Ninth Annual International Conference on International Computing Education Research', ICER '12, ACM, pp. 55–62.

Fawcett, T. (2006), 'An introduction to ROC analysis', *Pattern Recognition Letters* **27**(8), pp. 861–874.

Ferguson, R. & Buckingham Shum, S. (2011), Learning analytics to identify exploratory dialogue within synchronous text chat, *in* 'Proceedings of the 1st International Conference on Learning Analytics and Knowledge', LAK '11, ACM, New York, NY, USA, pp. 99–103.

Ferguson, R., Wei, Z., He, Y. & Buckingham Shum, S. (2013), An evaluation of learning analytics to identify exploratory dialogue in online discussions, *in* 'Proceedings of the Third International Conference on Learning Analytics and Knowledge', LAK '13, ACM, New York, NY, USA, pp. 85–93.

Filvà, D. A., García-Peñalvo, F. J. & Forment, M. A. (2014), Social network analysis approaches for social learning support, *in* 'Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality', TEEM '14, ACM, New York, NY, USA, pp. 269–274.

Frank, E., Hall, M. A. & Witten, I. H. (2016), The WEKA workbench, *in* 'Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques"', fourth edition edn, Morgan Kaufmann.

Fritz, J. (2011), 'Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers', *The Internet and Higher Education* **14**(2), pp. 89–97. Web mining and higher education: Introduction to the special issue.

Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. & Herrera, F. (2012), 'A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **42**(4), pp. 463–484.

Gama, J. (2010), *Knowledge Discovery from Data Streams*, 1st edn, Chapman & Hall/CRC.

Gama, J. a., Žliobaitė, I., Bifet, A., Pechenizkiy, M. & Bouchachia, A. (2014), 'A survey on concept drift adaptation', *ACM Comput. Surv.* **46**(4), pp. 1–37.

Gama, J., Medas, P., Castillo, G. & Rodrigues, P. (2004), Learning with drift detection, *in* 'Advances in Artificial Intelligence – SBIA 2004', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 286–295.

Garrison, D. R. & Kanuka, H. (2004), 'Blended learning: Uncovering its transformative potential in higher education', *The Internet and Higher Education* **7**(2), pp. 95–105.

Gerdes, H. & B, M. (1994), 'Emotional, social, and academic adjustment of college students: A longitudinal study of retention', *Journal of Counselling and Development* **72**(3), pp. 281–288.

Goggins, S. P., Galyen, K. & Laffey, J. (2010), Network analysis of trace data for the support of group work: Activity patterns in a completely online course, *in* 'Proceedings of the 16th ACM International Conference on Supporting Group Work', GROUP '10, ACM, New York, NY, USA, pp. 107–116.

Gray, G., McGuinness, C. & Owende, P. (2014), An application of classification models to predict learner progression in tertiary education, *in* 'Proceedings of IEEE International Advance Computing Conference (IACC)', pp. 549–554.

Gray, G., McGuinness, C., Owende, P. & Hofmann, M. (2016), 'Learning factor models of students at risk of failing in the early stage of tertiary education', *Journal of Learning Analytics* **3**(2), pp. 330–372.

Gunnarsson, B. L. & Alterman, R. (2012), Predicting failure: A case study in co-blogging, *in* 'Proceedings of the 2nd International Conference on Learning Analytics and Knowledge', LAK '12, ACM, New York, NY, USA, pp. 263–266.

Haig, T., Falkner, K. & Falkner, N. (2013), Visualisation of learning management system usage for detecting student behaviour patterns, *in* 'Proceedings of the Fifteenth Australasian Computing Education Conference - Volume 136', ACE '13, Australian Computer Society, Inc., Darlinghurst, Australia, pp. 107–115.

Hanley, J. A. & McNeil, B. J. (1982), 'The meaning and use of the area under a receiver operating characteristic (ROC) curve', *Radiology* **143**(1), pp. 29–36.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer Series in Statistics, second edition edn, Springer New York.

Hayes, D., Bernacki, M., Hong, W., Markle, J. & Voorhees, N. (2017), Using LMS data to provide early alerts to struggling students, *in* 'Proceedings of the 9th First Year Engineering Experience (FYEE) Conference', Daytona Beach, FL, USA.

He, J., Bailey, J., Rubinstein, B. I. P. & Zhang, R. (2015), Identifying at-risk students in massive open online courses, *in* 'Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence', AAAI'15, AAAI Press, pp. 1749–1755.

Hernández-García, A. & Conde, M. A. (2014), Dealing with complexity: Educational data and tools for learning analytics, *in* 'Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality', TEEM '14, ACM, New York, NY, USA, pp. 263–268.

Hira, Z. M. & Gillies, D. F. (2015), 'A review of feature selection and feature extraction methods applied on microarray data', *Advances in Bioinformatics* **2015**, pp. 1–13.

Hlosta, M., Zdrahal, Z. & Zendulka, J. (2017), Ouroboros: Early identification of at-risk students without models based on legacy data, *in* 'Proceedings of the Seventh International Learning Analytics & Knowledge Conference', LAK '17, ACM, New York, NY, USA, pp. 6–15.

Howard, E., Meehan, M. & Parnell, A. (2018), 'Contrasting prediction methods for early warning systems at undergraduate level', *The Internet and Higher Education* **37**, pp. 66–75.

Hu, Y.-H., Lo, C.-L. & Shih, S.-P. (2014), 'Developing early warning systems to predict students online learning performance', *Computers in Human Behavior* **36**, pp. 469–478.

Jackson, G.and Read, M. (2012), 'Connect 4 success: A proactive student identification and support program'.

Jayaprakash, S. M., Moody, E. W., Lauria, E. J. M., Regan, J. R. & Baron, J. D. (2014), 'Early alert of academically at-risk students: An open source analytics initiative', *Journal of Learning Analytics* **1**(1), pp. 6–47.

Jiang, S., Warschauer, M., E.Williams, A., O'Dowd, D. & Schenke, K. (2014), Predicting MOOC performance with week 1 behavior, *in* 'Proceedings of the 7th International Conference on Educational Data Mining', pp. 273–275.

Jishan, S. T., Rashu, R. I., Haque, N. & Rahman, R. M. (2015), 'Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique', *Decision Analytics* **2**(1), pp. 1–25.

Johnston, V. (1997), 'Why do first year students fail to progress to their second year? an academic staff perspective', *Paper presented at the British Educational Research Association Annual Conference* pp. pp. 11–14.

Kantardzic, M. (2011), *Ensemble Learning*, John Wiley and Sons, Ltd, chapter 8, pp. 235–248.

Kim, J. H., Park, Y., Song, J. & Jo, I.-H. (2014), Predicting students' learning achievement by using online learning patterns in blended learning environments: comparison of two cases on linear and non-linear model, *in* 'Proceedings of the 7th International Conference on Educational Data Mining', pp. 407–408.

Knutas, A., Ikonen, J. & Porras, J. (2013), Communication patterns in collaborative software engineering courses: A case for computer-supported collaboration, *in* 'Proceedings of the 13th Koli Calling International Conference on Computing Education Research', Koli Calling '13, ACM, New York, NY, USA, pp. 169–177.

Kohavi, R. & John, G. H. (1997), 'Wrappers for feature subset selection', *Artificial Intelligence* **97**(1), pp. 273–324.

Kolter, J. Z. & Maloof, M. A. (2003), Dynamic weighted majority: a new ensemble method for tracking concept drift, *in* 'Proceedings of the Third IEEE International Conference on Data Mining', pp. 123–130.

Lagus, J., Longi, K., Klami, A. & Hellas, A. (2018), 'Transfer-learning methods in programming course outcome prediction', *ACM Trans. Comput. Educ.* **18**(4), 19:1–19:18.

Laister, J. & Kober, S. (2002), Social aspects of collaborative learning in virtual learning environments, *in* 'Proceedings of the Networked Learning 2002 Conference'.

Larrabee Snderlund, A., Hughes, E. & Smith, J. (2018), 'The efficacy of learning analytics interventions in higher education: A systematic review', *British Journal of Educational Technology* **0**(0), pp. 1–25.

Leitner, P., Khalil, M. & Ebner, M. (2017), *Learning Analytics in Higher Education—A Literature Review*, Springer International Publishing, Cham, pp. 1–23.

Li, X., Wang, T. & Wang, H. (2017), Exploring n-gram features in clickstream data for MOOC learning achievement prediction, *in* 'Proceedings of the International Conference on Database Systems for Advanced Applications (DASFAA)', Springer International Publishing, Cham, pp. 328–339.

Liu, B. (2011), *Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data*, Springer-Verlag Berlin Heidelberg.

Loncar, M., Barrett, N. E. & Liu, G.-Z. (2014), 'Towards the refinement of forum and asynchronous online discussion in educational contexts worldwide: Trends and investigative approaches within a dominant research paradigm', *Computers & Education* **73**, pp. 93–110.

Lopez, M. I., Luna, J. M., Romero, C. & Ventura, S. (2012), Classification via clustering for predicting final marks based on student participation in forums, *in* 'Proceedings of the 5th International Conference on Educational Data Mining', pp. 148–151.

Macfadyen, L. P. & Dawson, S. (2010), 'Mining LMS data to develop an "early warning system" for educators: A proof of concept', *Computers & Education* **54**(2), pp. 588–599.

Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. & McClosky, D. (2014), The Stanford CoreNLP Natural Language Processing Toolkit, *in* 'Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations', Association for Computational Linguistics, pp. 55–60.

Manso-Vzquez, M. & Llamas-Nistal, M. (2015), 'A monitoring system to ease self-regulated learning processes', *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje* **10**(2), pp. 52–59.

Marbouti, F., Diefes-Dux, H. A. & Madhavan, K. (2016), 'Models for early prediction of at-risk students in a course using standards-based grading', *Computers & Education* **103**, pp. 1–15.

Mayilvaganan, M. & Kalpanadevi, D. (2014), Comparison of classification techniques for predicting the performance of students academic environment, *in* 'Proceedings of the 2014 International Conference on Communication and Network Technologies', IEEE, pp. 113–118.

McGill, T. J. & Klobas, J. E. (2009), 'A task-technology fit view of learning management system impact', *Computers & Education* **52**(2), pp. 496–508.

McKenzie, K. & Schweitzer, R. D. (2001), 'Who succeeds at university? factors predicting academic performance in first year australian university students', *Higher Education Research & Development* **20**(1), pp. 21–33.

Milliron, M. D., Malcolm, L. & Kil, D. (2014), 'Insight and action analytics: Three case studies to consider', *Research & Practice in Assessment* **9**, pp. 70–89.

Minaei-Bidgoli, B., Kashy, D., Kortemeyer, G. & Punch, W. (2003), Predicting student performance: an application of data mining methods with an educational web-based system, *in* 'Proceedings of the 33rd Annual Frontiers in Education, FIE 2003.', IEEE, pp. 13–18.

Mitchell, T. M. (1997), *Machine Learning*, McGraw-Hill Science Engineering Math.

Moodle (2017). viewed 16 May 2017.
URL: *https://moodle.com/about/*

Morris, L. V., Finnegan, C. & Wu, S.-S. (2005), 'Tracking student behavior, persistence, and achievement in online courses', *The Internet and Higher Education* **8**(3), pp. 221–231.

Mueen, A., Zafar, B. & Manzoor, U. (2016), 'Modeling and predicting students' academic performance using data mining techniques', *International Journal of Modern Education and Computer Science* **8**(11), pp. 36–42.

Muthukrishnan, S. M., Govindasamy, M. K. & Mustapha, M. N. (2017), 'Systematic mapping review on student's performance analysis using big data predictive model', *Journal of Fundamental and Applied Sciences* **9**(4S), pp. 730–758.

Na, K. S. & Tasir, Z. (2017), A systematic review of learning analytics intervention contributing to student success in online learning, *in* 'Proceedings of the 2017 International Conference on Learning and Teaching in Computing and Engineering (LaTICE)', pp. 62–68.

Natek, S. & Zwilling, M. (2014), 'Student data mining solutionknowledge management system related to higher education institutions', *Expert Systems with Applications* **41**(14), pp. 6400–6407.

Osmanbegovic, E. & Suljic, M. (2012), 'Data mining approach for predicting student performance', *Journal of Economics and Business/Economic Review* **10**(1), pp. 3–12.

Pantages, T. J. & Creedon, C. F. (1978), 'Studies of college attrition: 1950-1975', *Review of Educational Research* **48**(1), pp. 49–101.

Pardo, A., Mirriahi, N., Martinez-Maldonado, R., Jovanovic, J., Dawson, S. & Gašević, D. (2016), Generating actionable predictive models of academic performance, *in* 'Proceedings of the Sixth International Conference on Learning Analytics & Knowledge', LAK '16, ACM, New York, NY, USA, pp. 474–478.

Polikar, R. (2006), 'Ensemble based systems in decision making', *IEEE Circuits and Systems Magazine* **6**(3), pp. 21–45.

Polikar, R. (2012), *Ensemble Learning*, Springer US, Boston, MA, pp. 1–34.

Rabbany k., R., Takaffoli, M. & Zaïane, O. R. (2012), 'Social network analysis and mining to support the assessment of on-line student participation', *SIGKDD Explor. Newsl.* **13**(2), pp. 20–29.

Rahman, N. & Dron, J. (2012), Challenges and opportunities for learning analytics when formal teaching meets social spaces, *in* 'Proceedings of the 2nd International Conference on Learning Analytics and Knowledge', LAK '12, ACM, New York, NY, USA, pp. 54–58.

Raza, H., Prasad, G. & Li, Y. (2015), 'EWMA model based shift-detection methods for detecting covariate shifts in non-stationary environments', *Pattern Recognition* **48**(3), pp. 659–669.

Robinson, C., Yeomans, M., Reich, J., Hulleman, C. & Gehlbach, H. (2016), Forecasting student achievement in MOOCs with natural language processing, *in* 'Proceedings of the Sixth International Conference on Learning Analytics and Knowledge', LAK '16, ACM, New York, NY, USA, pp. 383–387.

Rogers, T., Colvin, C. & Chiera, B. (2014), Modest analytics: Using the index method to identify students at risk of failure, *in* 'Proceedings of the Fourth International Conference on Learning Analytics And Knowledge', LAK '14, ACM, New York, NY, USA, pp. 118–122.

Romero, C., Lopez, M.-I., Luna, J.-M. & Ventura, S. (2013), 'Predicting students' final performance from participation in on-line discussion forums', *Computers & Education* **68**, pp. 458–472.

Sclater, N., Peasgood, A. & Mullan, J. (2016), 'Learning Analytics in Higher Education. A review of UK and international practice. Full report'. Retrieved from.

Shalizi, C. R., Jacobs, A. Z., Klinkner, K. L. & Clauset, A. (2011), 'Adapting to non-stationarity with growing expert ensembles', https://arxiv.org/abs/1103.0949.

Shaw, R.-S. (2012), 'A study of the relationships among learning styles, participation types, and performance in programming language learning supported by online forums', *Computers & Education* **58**(1), pp. 111–120.

Shelton, B. E., Hung, J.-L. & Baughman, S. (2016), 'Online graduate teacher education: Establishing an EKG for student success intervention', *Technology, Knowledge and Learning* **21**(1), pp. 21–32.

Shelton, B. E., Hung, J.-L. & Lowenthal, P. R. (2017), 'Predicting student success by modeling student interaction in asynchronous online courses', *Distance Education* **38**(1), pp. 59–69.

Siemens, G. (2011), 'The 1st international conference on learning analytics and knowledge'. viewed 20 November 2017.
URL: *https://tekri.athabascau.ca/analytics/*

Siemens, G. & Baker, R. S. J. d. (2012), Learning analytics and educational data mining: Towards communication and collaboration, *in* 'Proceedings of the 2Nd International Conference on Learning Analytics and Knowledge', LAK '12, ACM, New York, NY, USA, pp. 252–254.

Smith, V. C., Lange, A. & Huston, D. R. (2012), 'Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses', *Journal of Asynchronous Learning Networks* **16**(3), pp. 51–61.

Soares, S. G. & Araujo, R. (2015), 'An on-line weighted ensemble of regressor models to handle concept drifts', *Engineering Applications of Artificial Intelligence* **37**, pp. 392–406.

Sorour, S. E., Abd El Rahman, S., Kahouf, S. A. & Mine, T. (2016), Understandable prediction models of student performance using an attribute dictionary, *in* 'Proceedings of the 15th International Conference on Advances in Web-Based Learning (ICWL)', Springer International Publishing, Cham, pp. 161–171.

Spady, W. G. (1970), 'Dropouts from higher education: An interdisciplinary review and synthesis', *Interchange* **1**(1), pp. 64–85.

Stanford CoreNLP Natural language software (2015). viewed 12 March 2018.
URL: *https://stanfordnlp.github.io/CoreNLP/*

Street, W. N. & Kim, Y. (2001), A streaming ensemble algorithm (SEA) for large-scale classification, *in* 'Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '01, ACM, New York, NY, USA, pp. 377–382.

Terenzini, P. T. & Pascarella, E. T. (1978), 'The relation of students' precollege characteristics and freshman year experience to voluntary attrition', *Research in Higher Education* **9**(4), pp. 347–366.

Thammasiri, D., Delen, D., Meesad, P. & Kasap, N. (2014), 'A critical assessment of imbalanced class distribution problem: The case of predicting freshmen student attrition', *Expert Systems with Applications* **41**(2), pp. 321–330.

The Online Plain Text English Dictionary (OPTED) (2000).
URL: *https://www.mso.anu.edu.au/~ralph/OPTED/*

Tinto, V. (1975), 'Dropout from higher education: A theoretical synthesis of recent research', *Review of Educational Research* **45**(1), pp. 89–125.

Tucker, C. S., Pursel, B. & Divinsky, A. (2014), 'Mining student-generated textual data in MOOCs and quantifying their effects on student performance and learning outcomes', *Computers in Education Journal* **5**(4), pp. 84–95.

Tukey, J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley.

Villagra-Arnedo, C. J., Gallego-Duran, F. J., Llorens-Largo, F., Compan-Rosique, P., Satorre-Cuerda, R. & Molina-Carmona, R. (2017), 'Improving the expressiveness of black-box models for predicting student performance', *Computers in Human Behavior* **72**, pp. 621–631.

Vrasidas, C. & McIsaac, M. S. (1999), 'Factors influencing interaction in an online course', *American Journal of Distance Education* **13**(3), pp. 22–36.

Wang, A. Y. & Newlin, M. H. (2002), 'Predictors of performance in the virtual classroom: Identifying and helping at-risk cyber-students', *T H E Journal* **29**(10), pp. 21–25.

Wang, A. Y., Newlin, M. H. & Tucker, T. L. (2001), 'A discourse analysis of online classroom chats: Predictors of cyber-student performance', *Teaching of Psychology* **28**(3), pp. 222–226.

Wang, H., Fan, W., Yu, P. S. & Han, J. (2003), Mining concept-drifting data streams using ensemble classifiers, *in* 'Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining', KDD '03, ACM, New York, NY, USA, pp. 226–235.

Wang, H., Fan, W., Yu, P. S. & Han, J. M.-B. (2006), Early drift detection method, *in* 'Proceedings of the Fourth International Workshop on Knowledge Discovery from Data Streams'.

Wen, M., Yang, D. & Rose, C. (2014*a*), Linguistic reflections of student engagement in massive open online courses, *in* 'Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media', AAAI Press, Ann Arbor, Michigan, USA, pp. 525–534.

Wen, M., Yang, D. & Rose, C. (2014*b*), 'Sentiment analysis in MOOC discussion forums: What does it tell us?', *Educational Data Mining* pp. 130–137.

Wilson, S. (2005), 'Improving retention and success: A case study approach for practical results', *Journal of College Student Retention: Research, Theory & Practice* **7**(3), pp. 245–261.

Wise, A. F. (2014), Designing pedagogical interventions to support student use of learning analytics, *in* 'Proceedings of the Fourth International Conference on Learning Analytics And Knowledge', LAK '14, ACM, New York, NY, USA, pp. 203–211.

Wise, A. F., Zhao, Y. & Hausknecht, S. N. (2014), 'Learning analytics for online discussions: Embedded and extracted approaches', *Journal of Learning Analytics* **1**(2), pp. 48–71.

Wolff, A., Zdrahal, Z., Nikolov, A. & Pantucek, M. (2013), Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment, *in* 'Proceedings of the Third International Conference on Learning Analytics and Knowledge', LAK '13, ACM, New York, NY, USA, pp. 145–149.

Wolfgang, G. & Hendrik, D. (2012), 'Translating learning into numbers: A generic framework for learning analytics', *Journal of Educational Technology & Society* **15**(3), pp. 42–57.

Wong, B. T. M. (2017), 'Learning analytics in higher education: an analysis of case studies', *Asian Association of Open Universities Journal* **12**(1), pp. 21–40.

Wong, B. T. M. & Li, K. C. (2018), Learning analytics intervention: A review of case studies, *in* 'Proceedings of the 2018 International Symposium on Educational Technology (ISET)', pp. 178–182.

Xing, W., Guo, R., Petakovic, E. & Goggins, S. (2015), 'Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory', *Computers in Human Behavior* **47**, pp. 168–181.

Xu, C., Zheng, Y., Hu, H. & Li, Y. (2016), 'Exploring individual contributions during online collaborative discussions', *Journal of Computers in Education* **3**(4), pp. 395–411.

Yang, D., Wen, M., Howley, I., Kraut, R. & Rose, C. (2015), Exploring the effect of confusion in discussion forums of massive open online courses, *in* 'Proceedings of the Second (2015) ACM Conference on Learning @ Scale', L@S '15, ACM, New York, NY, USA, pp. 121–130.

Yukselturk, E. (2010), 'An investigation of factors affecting student participation level in an online discussion forum', *Turkish Online Journal of Educational Technology - TOJET,* **9**(2), pp. 24–32.

Zliobaite, I. & Gabrys, B. (2014), 'Adaptive preprocessing for streaming data', *IEEE Transactions on Knowledge and Data Engineering* **26**(2), pp. 309–321.