UMBC

DISSERTATION APPROVAL SHEET

Title of Dissertation: PROFILING AND MODELING STUDENT LEARNING BEHAVIORS & OUTCOMES FROM DIGITAL LEARNING ENVIRONMENTS

Name of Candidate: Varun Mandalapu

Doctor of Philosophy, 2021

Graduate Program: Information Systems

Dissertation and Abstract Approved:

Zhiyuan Chen

Zhiyuan Chen
Professor
Information Systems
9/5/2021 | 5:33:05 PM EDT

Dr. Jiaqi Gong

Dr. Jiaqi Gong
Associate Professor (University of Alabama)
Computer Science
9/6/2021 | 7:39:09 PM EDT

NOTE: *The Approval Sheet with the original signature must accompany the thesis or dissertation. No terminal punctuation is to be used.

ABSTRACT

Title of Document:	Profiling and Modeling Student Learning Behaviors & Outcomes from Digital Learning Environments
	Varun Mandalapu, Doctor of Philosophy, 2021
Directed By:	Dr. Zhiyuan Chen, Professor
	Department of Information Systems, UMBC
	Dr. Jiaqi Gong, Associate Professor
	Department of Computer Science, University of Alabama

Educational data mining focuses on exploring increasingly large-scale data from educational settings, such as Learning Management Systems (LMS), and developing computational methods to understand students' behaviors and learning settings better. There has been a multitude of research dedicated to studying the student learning process, leading to multiple commonly cited frameworks and theories that characterize students' learning behaviors. However, the most recent focus of interest argued that developing efficient models with actionable suggestions that help understand student learning behaviors and contribute to their academic outcomes is needed. Existing studies investigated various individual, emotional, and social factors related to student learning behaviors by analyzing fine-grained data samples collected by LMS at the student level but fail to incorporate the dynamics of learning behaviors and fail to examine the whole image of the student learning lives. For instance, most research in student learning behavior focuses on specific courses and related academic performance but did not consider that students always take multiple courses simultaneously. Therefore, the implications and

knowledge from the static understanding of students' learning behavior in an isolated specific course may be limited to generate actionable strategies to help students. This dissertation was motivated to explore large-scale data, especially for examining the learning behavior's dynamics and developing student-centric models. The resulting knowledge has been recognized by publications in top-level international conferences in educational data mining and artificial intelligence in education.

The first research attempt of my research introduced a new computational method based on psychological theories in affect dynamics to track dynamic student behaviors and developed a novel explanation method based on Local Interpretable Model-Agnostic Explanations (LIME). The second research attempt focused on developing computational models at the student level to predict their academic performance based on LMS data at the University of Maryland, Baltimore County. The findings from this work showed that student login volume and their prior performance significantly impact student performance. Additionally, this research focused on exploring causal relationships between student LMS behaviors and their academic performance. The causal analysis strengthened our findings in computational modeling by showing a significant cause-and-effect relationship between student login behaviors and their academic performance. The conclusions from this work will empower intervention techniques that improve student emotion regulation capabilities. The student-centric models developed in this study reported the positive impact of student login behaviors on their academic performance. This understanding will enable LMS developers and school administrators to design and develop interactive systems that deliver course content effectively.

PROFILING AND MODELING STUDENT LEARNING BEHAVIORS & OUTCOMES FROM DIGITAL LEARNING ENVIRONMENTS

By

Varun Mandalapu

Dissertation submitted to the Faculty of Graduate School of the

University of Maryland, Baltimore County, in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

2021

© Copyright by

Varun Mandalapu

2021

Dedication

I dedicate my dissertation work to my family and friends. A special feeling of gratitude to my loving parents, Anjaneyulu and Padmavathi whose words of encouragement and push for tenacity ring in my ears. My wife Sharmila, her parents and my daughter Siri have never left my side and supported me throughout my PhD journey.

I also dedicate this dissertation to my many friends who have supported me throughout the process.

Acknowledgements

I owe my deepest gratitude to all those people who have made this dissertation possible, and because of whom my graduate experience has been one that I will cherish forever.

First and foremost, I would like to express my sincere gratitude to my graduate advisor Dr. Jiaqi Gong. I thank him for his constant support, guidance, and continued belief in me throughout this thesis work. Throughout my studies, he always has the highest priority for his students, which is evident in his leadership success. I am proud that I have worked and learned a lot from such a prolific scientist and researcher. I would like to acknowledge my dissertation co-advisors Dr. Zhiyuan Chen, and Dr. Lujie Karen Chen with whose support I gained new knowledge and guidance to conduct my research effectively and efficiently. I am also grateful to my committee members Dr. Karuna Pande Joshi and Dr. Charissa Cheah (UMBC Psychology) for their comments and guidance while holding me to a high research standard. I want to thank Dr. John Fritz, Dr. Thomas Penniston and Kevin Joseph from Division of Information Technology at UMBC, for their dedication and commitment to help with knowledge and access to data related to my research.

I must thank all of the Information Systems department professors with whom I served as teaching assistant over the last years for showing me what it means to be dedicated and helping me excel in my university teaching skills, each in their unique way. Each of you have given your time, energy, and expertise, and I am richer for it: Dr. Jiaqi Gong, Prakash Bandaru and Dr. Zhiyuan Chen. During my stay at UMBC, I made numerous friends who have helped me stay motivated through these years and have made my journey joyful. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value the friendship of my current and alumnus fellow lab mates, Hee-ra Lee, Xishi Zhu, Dae-young Kim, Stephan Kaputsos and Truc Tran. I also thank my fellow PhD student friends Sreenivasan Ramasamy Ramamurthy, Lavanya Elluri and Neha Singh for their collaborations and support. I sincerely appreciate their help and belief in me.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

Researchers in education must identify and measure the factors that influence student motivation and engagement while learning to understand productive learning behaviors. Studies in psychology categorized these factors into four major categories: individual, school-related, family-related, and social [1][2][3][4]. The individual factors of a student are defined based on their skills and characteristics. These skills focus on student self-regulation, self-efficacy, and emotion regulation capabilities during learning, whereas their characteristics are related to beliefs, values of education, ethnic identity, and orientation. While students' skills directly influence their cognitive abilities, their characteristics influence their perception of learning [5]. The second category that influences student learning is based on school-related factors. This category consists of factors that are experienced in a school environment, such as institutional practices, teacher support, and student academic achievement [6]. The third category focuses on family-related factors that include student family background, family obligations, and family support [2]. The final category related to social factors includes social and cultural constructs of students like peer support, neighborhood situation, and discrimination in the societal context [3].

Existing research in educational data mining investigates the factors mentioned above by modeling large-scale data captured by digital learning environments. Student individual factors were analyzed by developing models that predict patterns related to emotion, self-regulation, and efficacy. These models combine fine-grained student interaction data captured by LMS at different levels (e.g., course level and assignment level) with self-assessments and field observations [7][24]. Similarly, the influence of social and family-related factors on student learning behaviors are being studied by integrating student demographic data (e.g., gender, income level, locality, and family factors) with their outcomes captured by LMS [8][9]. While all these categories impact student learning behaviors, this current study focuses on profiling and modeling student's academic emotions, learning behaviors, and self-regulation strategies based on time characteristics to support student academic achievement. It is essential to understand the role of emotions and learning strategies in an academic setting to develop these student profiles. Emotions are ubiquitous in an educational environment [10]. For example, when students study learning material, they may enjoy learning or getting bored or frustrated. This emotional change is because of never-ending obstacles or experience a flow of positive emotions that makes them forget the time based on their learning goals and content present in the material. Furthermore, these emotions will affect student motivation, effort, and learning strategies like time management [11].

Empirical findings show that students experience a wide variety of emotions while attending class, doing assignments, and taking exams and tests [12][13]. A study conducted by Pekrun and colleagues on emotions experienced by students at universities found that anxiety, boredom, frustration, anger, enjoyment, and concentration are frequently encountered in an academic setting [53]. Until recently, the emotions mentioned above were understudied except for test anxiety [14][15]. In the past decade, however, there has been a growing literature on academic emotions' impact on student learning, achievement, health, and personality development. These studies focused on the way to model and profile student academic emotions to develop emotional intelligence strategies that support student's academic achievement [16][17][18]. However, these studies also encounter challenges in validating the developed models as they are primarily dependent on student self-assessments prone to self-selection bias and biased memories.

To resolve the challenges mentioned above, research in educational data mining and learning analytics focused on developing quantitative affect detection from LMS data to model student emotional behaviors while learning [19][24]. The current affect detection methods are both sensor-based and sensor-free. The sensor-based affect detection methods deploy physical sensors like video cameras, posture detection systems, and eye trackers in student learning environments [19][20][21]. These systems will then capture student physical attributes like gaze, head movements, and posture while interacting with their learning material. This data collected by sensors is then used to model their affect states in conjunction with learning tasks. However, these systems suffer from high cost, scalability, and privacy issues. To avoid these issues, research in educational data mining focused on sensor-free affect detection systems.

Sensor-free affect detection systems use data captured by digital learning environments and quantitative field observations to develop models that predict student's affective or emotional states while learning [24]. Baker, Rodrigo, and Ocumpaugh developed a quantitative field observation method called Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP). The BROMP protocol observes students during their interaction with digital learning environments and codes their affective states such as concentration, confusion, boredom, and frustration [66]. Later studies developed novel affect detection methods by appending the affect labels from BROMP with student interaction, knowledge, and timing data captured by digital learning environments [22][24][72]. These studies use advanced machine learning models to develop automated sensor-free affect detection systems. While these studies developed better detection models, their performance is still above chance and lacks explanations about the impact of student affect on student outcomes. One primary reason for this is the development of models based on individual affect states instead of dynamics. Studies in psychology argue that modeling students' affect dynamics and chronometry is vital to understanding the impact of learning emotions [130][131]. This current study develops high-performing sensor-free affect detection methods by profiling student affect from fine-grained data captured by digital learning environments in conjunction with quantitative field observations. The methods proposed in this current study also develop a novel model explanation method that helps understand the impact of affect, knowledge, and interaction features captured by digital learning environments and student affect states on student outcomes.

While academic emotions significantly influence student learning behavior, research in education and psychology argues that students' learning models should incorporate socioenvironmental and self-regulation strategies [23][25]. According to Winne and Jamieson-Noel [219], students manage their learning process by setting goals, defining tasks, and choosing strategies required to achieve these goals. Hence, the capacity of a student to plan and schedule time effectively is an essential part of self-regulation that contributes to student academic success. For example, students who organize their study time better perform well and get better grades, whereas students that exhibit procrastination behavior have below-average grades in their academic careers. Studies also inform that time management skills are linked to academic motivation contributing to student success [220][221]. While it is evident that managing time is an essential aspect of student learning, it is not yet understood what kind of time management strategies students use while learning. Existing studies focus on studying time management strategies based on self-assessment questionnaires linked to student motivation. A survey by Britton and Tesser [220] developed a time management questionnaire that focuses on student schedules and their learning goals.

Similarly, Macan et al. [222] developed another questionnaire to study stress-coping theory based on time management. While these studies provide valuable insights about time management and student learning, it is hard to validate and study them in real-time. The current study profiles student learning behaviors and time characteristics based on their login patterns, course content access, and time spent on different learning tasks scheduled at the student level. This study develops student-centric models based on login behavior data captured by LMS to predict student academic performance.

In addition to login related features, the chronotypes of a student influence their learning outcomes. A recent study focused on identifying the relationship between the social learning environment, chronotypes, and student engagement found a positive correlation [26][27]. These studies also informed that students' choice of learning environment and time of learning depends on their social and economic aspects. For example, a student coming from low-income households works during the school year and continues their education. Due to the limited availability of time and tight work schedules, students who work and study during the academic year choose informal learning environments and times compared to full-time students living on campus. However, there is little empirical evidence on the role of informal learning places like online environments on student engagement and success. Therefore, the proposed study will include a chronotype analysis based on student demographics to understand the LMS access times and correlate with the student performance to understand the impact of learning environment on student engagement.

By profiling and modeling factors that support students' learning process, this dissertation aims to provide broader impacts in the research of student affect and learning strategies that lead to actionable insights into these constructs.

1.1 Research Questions

This study is divided into three parts, with each part focusing on one research question. The three research questions are listed below.

- Can the profiles of students’ affect (e.g., transitions among the affective states) be predicted based on user interactions recorded by digital learning environments that support understanding of student’s affect dynamics in learning?
- Can the patterns be extracted and prioritized from student interaction profiles for predicting student learning and career outcomes? This question also focuses on understanding the impact of affective dimensions on student outcomes such as career choices using explainable AI.
- How the mechanisms of self-regulation strategies, such as time characteristics and learning behaviors, influencing student motivation and learning process, are encoded in the data recorded by the learning management systems? We used the Blackboard system data at a minority-serving institute as a concrete example.

1.2 Research Objective & Impacts

This research aims to arrive at a richer understanding of how the students' learning process can be predicted and tracked by using their interactions with digital learning environments. This study primarily focuses on affective dimensions and environmental constructs that impact student learning abilities. While existing research has examined the high-level emotional impacts on student learning [28][29][30][31], studies on deep academic emotions have not garnered a significant amount of attention. The overarching objective of this dissertation is a contribution to the development of novel affect detection methods and provide insights into the influence of affective dimensions and various environmental constructs on student learning outcomes.

This research will assist in developing computational models that predict student emotional behaviors that help educational systems to develop methods that help students achieve their academic and career goals. As the implementation of information systems that capture vast amounts of data is increasing daily in educational settings, this research will provide directions to analyze and infer in real-time with a proper context. These systems will also help deliver interventions in real-time that helps students manage their self-learning capability and support them to be lifelong learners.

This digital learning-based empowerment tool to students' self-regulated learning by improving their persistence could significantly impact students' cognitive engagement, thereby changing students as lifelong learners and increasing the efficacy of higher education. This study will help understand the student's ability of self-regulated learning to keep up with the explosive growth of knowledge and skills in their career and to retool into a new career after their previous one runs its course. Besides, this study will provide an additional dimension of the science of learning, specifically, the fundamental understanding of human learning behavior, measured by the learning systems, such as the contextual factors (e.g., emotions, engagement, social events) related to self-regulation.

This understanding will help us revisit previous knowledge in learning based on self-reported measures and improve the reproducibility of research in teaching and learning.

1.3 Contributions

The overarching objective of this dissertation is a contribution to the development of efficient data-driven models from LMS that profile student learning behaviors based on individual and socio-environmental constructs. Recent studies in education identified that student emotional behaviors significantly impact their learning constructs like motivation and self-regulation. Through the development of novel computational modeling techniques that help track student affect while learning, this proposed study helps facilitators develop intervention techniques that guide student emotions in a way that leads to their academic success. The scalable approach to explore the emotional lives of students could have a significant impact on the understanding of contextual and dynamic processes of learning behavior that contribute to their learning outcomes.

The previous studies in education and computer science focused on developing models that predict student outcomes from digital learning environments failed to explain the underlying reasons and impact of different dimensions on such outcomes. This study develops a novel model explanation method that helps researchers understand the importance of different dimensions extracted from digital learning environments in predicting student outcomes. This understanding of the role of multiple dimensions in student outcomes will help educational facilitators take steps that focus on improving student engagement and validating them based on domain experts' inputs.

Finally, this work tries to profile students based on their self-regulation strategies like time management to understand the learning tactics used by students to improve their academic performance. When combined with demographics and learning environments, these strategies will provide a richer understanding of the impact of various constructs on student learning practices. The research results from this study will provide a strong foundation for adopting these methods in the broader population.

1.4 Dissertation Outline

Chapter 1:

In this chapter, a background of the problem space is provided and details about the specific problem areas this dissertation addresses. The research objective and the broader impacts of this research are discussed in section 1.1. Finally, the three major research questions focused on in this dissertation are listed in section 1.2.

Chapter 2:

Chapter 2 reviews the areas of research related to the problems defined in this dissertation. This section mainly focuses on research related to the student learning process and relevant learning theories. This chapter also covers current digital learning environments, the affective dimensions of student learning, learning analytics, and relationships.

Chapter 3:

Chapters 3 in this study focuses on novel affect dynamics methods that help capture student's affect transitions from digital learning environments. This chapter will discuss the importance of finding affect transitions and their ability to help improve the student learning process.

Chapter 4:

Chapters 4 details a novel method to predict student outcomes by modeling their interactions and affect dimensions. It also details a novel method of understanding predictions given by machine learning models using concepts from explainable AI.

Chapter 5:

Chapters 5 in this study focuses on developing student-centric models to predict their end of term GPA based on login variables captured by Learning Management Systems data. This chapter also discusses the importance of different login variables on student academic performance and explores their variations based on student demographics.

Chapter 6:

The final chapter in this study focuses on studying student time characteristics based on their login information on LMS system. This chapter also explores causal relationships between student login behaviors, chronotypes and academic performance.

Chapter 2: Related Work

Over the past decades, theories and research findings from educational psychology, learning theory, student development, curriculum design, and instructional design have been developed and examined across various aspects: behavior, brain, physiology, and psychology. Numerous researchers have investigated the relations between these theories and research findings to develop a unified theoretical framework and actionable intervention strategies. However, these methodologies needed to handle the network and systems approaches for existing models and theories are open issues educational data mining domain.

2.1 Learning Theories

A recent review related to the learning process, issues of intelligence thinking, curriculum design, and human development listed 111 theories from 1924 to 2010. The authors of this study discussed several theories regarding the principles of learning and practical instructional strategies and provided a unified explanation of the phenomenon of learning [32]. The learner-centered explanation emphasized the learner's effort to make meaning of an external world, inner self, and the relation between the two. Specifically, an individual's value orientation, perceived needs, cognitive capabilities, intrinsic motivation, and flow of emotions trigger, drive and direct their meaning-making processes and efforts [32].

The multi-faceted learning process has been investigated across various aspects: cognitive, motivational, instructional, and institutional. Most recently, researchers argued that more attention should be paid to the affective process of learning because of the nontrivial impacts of emotions on learning [33][34][35][36]. Despite strong evidence for the positive effects of positive emotions, no clear rules such as positive emotions foster and negative emotions deter learning, as mentioned in this study. The results from multiple studies are sometimes equivocal because of the limited temporal understanding of the emotions measured using a self-reported questionnaire. Exploring a high-resolution assessment of emotional episodes "at the moment" would enhance our understanding of the association between emotions and learning. Another critical theory regarding student's lives developed by Lazarus pertains to stress and coping [37][38][39]. Specifically, the stress induced by the underlying barriers and obstacles might be more intense and detrimental from students performing at a low level.

2.2 Digital Learning Environment

With advances in information technology, current learning environments such as the learning management system [40][41][42][43][44], intelligent tutoring systems, and Massive Open Online Courses (MOOC) [45][46][47] has attracted much attention from researchers, institutions, and stakeholders. However, researchers argue that these systems are not the solution to every problem in education. Specifically, EDUCAUSE in 2015 envisioned the Next Generation Digital Learning Environment (NGDLE) to improve our society through education [48]. The five themes of the NGDLE include interoperability and integration, analytics, personalization, collaboration, advising and learning assessment, and accessibility and universal design.

Current research on the learning environment is expanding, multidisciplinary in nature, and embodies the five themes mentioned in NGDLE. Most of the research attempts to leverage artificial intelligence for generating smart components of the learning environment to achieve these five themes [49][50][51]. While most of the researchers focus on NGDLE, a most recent literature review revealed that there is currently a less comprehensive understanding of the theme of personalization, such as learner psychology [52]. Understanding the learning mechanism and student learning behavior through educational data mining from multimodal, multiscale, and heterogeneous data captured within the learning environment remains an understudied issue but could achieve personalization for students.

2.3 Psychological Connections to Learning Theories

Pekrun defines a complex, markedly symmetrical taxonomy of emotions, including achievement emotions, epistemic emotions, and social emotions [53][54]. The ongoing validation of an instrument has parallelled the development of academic emotion theory, the Achievement Emotions Questionnaire (AEQ), which has generated variants for use in different age groups, subject areas, and languages in addition to the original English. The multiple axes and domains posited by academic emotion theory allow for fine-grained classification of the vast vocabulary of emotion words relating to schoolwork that may be heard by a practicing school counselor or educational psychologist. A counselor aims to understand a learner's natural language, not to spend time splitting hairs to fit the learner's self-expression into a scientifically verifiable taxonomy. A most recent study reviewed affective computing in educational research and reported that most affective computing studies on learning adopted the AEQ instrument.

Lazarus and his colleagues have developed the stress and coping theory over many years [37][38][39][55]. Within his conceptual framework, stress is defined as a relationship between the person and the environment that the person appraises as relevant to his or her well-being and in which the person's resources are taxed or exceeded. Mainly, this definition of stress accurately reflects the situation the underrepresented students encountered [56][57][58][59][60][61]. Numerous researches have reported that minority students had a higher tendency to experience stress than their counterparts, and the perceived stress profoundly impair the relationship between the students and the environment, which impair their STEM persistence [62][63][64]. Therefore, according to the theory of stress and coping, examining the underrepresented group's stress status and coping strategies would generate a deeper understanding of the students' learning behavior, thus developing effective intervention programs.

Social Cognitive Career Theory (SCCT) develops a conceptual model for implementing factor analysis to understand the barriers and obstacles to the underrepresented group [56][65]. Specifically, it proposes that self-efficacy beliefs are formed via four types of information: performance accomplishments (e.g., personal mastery experiences or past successes), vicarious learning (e.g., observing the explicit behaviors of others, such as role models), social persuasion (e.g., verbal encouragement), and affective/emotional arousal experienced while completing a task (e.g., low anxiety, relaxed state) [56]. Notably, the theory considers several pathways through which individual differences (e.g., gender, race, and personality) affect the academic and career development process.

These educational theories provide insights regarding the link between the emotional aspects of student life and the learning outcomes. However, the context and dynamics of the conceptual link are not well understood yet. Therefore, new investment in improving the resolution of emotion episodes and monitoring the context of emotions is needed for future research. The insights derived from the context and dynamics of the emotional lives of the students are likely to significantly impact the scientific progress in exploring the relationship between emotion and learning, thus providing more in-depth practice guidelines for intervention programs for students.

2.4 Affective Dimensions of Learning

Scholars in educational psychology reviewed the positive and negative impacts of affective states on student cognitive processes [66][67]. These studies state that positive affect like concentration impacts cognitive processes and improves their learning capabilities, while negative affect states will hinder these abilities [66]. The four most significant affect states that impact individual student learning includes engaged concentration, confusion, boredom, and frustration. Studies focusing on understanding the various implications of student's affect states observed during their middle school years found correlations with major selection in college, career choice, and performance [68][69]. Educational technologies like ITS, MOOC courses that can capture student interactions and knowledge levels, when combined with human quantitative field observations (BROMP), reinforced predictive models in affect state detection [70].

The increased interest in leveraging affective computing to assess the student's affective states has generated many successful approaches to building accurate affect detectors. Affect detectors solely rely on the learning data captured by the computer-based learning systems, named “Sensor-free affect detectors” [71]. Sensor-free affect detectors have been investigated for many years, and the models learned from previous studies have been verified and validated in the diverse dataset [71][72]. Currently, the best sensor-free affect detector that depends on extracting knowledge states from student interaction with

ITS using Deep learning methods has demonstrated better than random chance performance to detect engaged student concentration, confusion, frustration, and boredom solely from student's log data [71]. The datasets used in these studies include the cognitive tutor dataset and the ASSISTments dataset. The detectors (predictive models) built from these datasets helped various researchers study the influences of students' affective states on various aspects.

Meanwhile, much effort, such as data cleaning for missing skills and wrong answers, new training models leveraging the deep learning techniques, have been invested in increasing the accuracy of sensor-free affect detectors and eventually using them to drive intervention. However, little has been reported to develop affect detectors leveraging existing psychological theory regarding affect dynamics. For instance, researchers in affect styles have demonstrated that the temporal features of affect (e.g., affect change) include richer information than affective states and expose the individual progress ineffective regulation. This study incorporates the affect concepts from psychology for building computational models to develop a better affect tracking system.

2.5 Learning Analytics Research

LMS systems have the ability to capture large data streams related to user interactions through which administrators and instructors can develop methods to improve the learning experience. The collection, analysis, and reporting of data about learning activities on web-enabled learning platforms to assess student academic progress, performance prediction, and potential issues that need attention is the central proposition of emerging fields like learning analytics and educational data mining [73][74]. Outcomes derived from learning analytics aim to gain insights about student learning behaviors, real-time information about institutional practices, and support designing personalized courses in CMS. Although there are huge data stores in universities and colleges that can be used to make data-driven decisions to support optimal use of both pedagogical and economic resources, there has been the minimal application of this data in higher education [75].

An LMS system records student interaction details related to logins, number of posts written on discussion threads, time spent on lecture materials, total downloads, etc., in their log files. These logs can be analyzed to generate reports that help teachers to observe student progress at a granular level. Once there are enough student records collected in LMS, they can be used to develop computational models to predict future student performances. Multiple works in EDM and LA studied the relation between the usage of LMS and student academic achievements. Vengroff and Bourbeau's [76] study showed evidence that providing additional material in LMS benefited students at the undergraduate level. They also conclude that students who used LMS regularly did better in exams than their peers who have minimal interactions. In their research, Dutt and Ismail [77] observed that tracking resources students interact with on LMS supports developing new strategies that make learning easier and enhance learner progress. Their work also focused on analyzing thresholds related to student interaction features like self-assessment tests, time spent on exercises, discussion forums, and performance outcomes. Another study by Lust et al. [78] explored the usage variations in different tools used by students on LMS, such as time on web-link, time on web-lectures, time on a quiz, time on feedback, postings on discussion board, and messages read. The results from this study heavily contributed to the development of adaptive and innovative recommendation systems. In their work, Hung and Zhang [79] also found patterns based on six indices that represent student effort: Frequency of accessing the course material, number of LMS logins, total interactions in discussion threads, number of synchronous discussions, number of posts read, and final grades in a course.

While exploring a link between student online activity on LMS and their grades, Dawson et al. [75] observed a significant difference in the number of online sessions accessed, total time spent, and the number of posts in discussion forums between high and low performing students. Another study by Damainov et al. [80] developed a multinomial logistic regression model based on time spent in LMS. This study found a significant relationship between student time spent and grades, especially in students who attained lower grades between D and B. Instead of using time spent online, other works focused on the frequency of course material access within LMS. A study by Baugher et al. [81] found that regularity in student hits is a reliable predictor of student performance compared to the total number of hits. In their study, Chancery and Haque analyzed student interaction logs of 112 undergraduate students. They found students with low LMS access rates obtained lower grades than their peers with higher access rates. This study was complemented by Biktimirovan and Klassen [82] that reported a strong relationship between student hit consistency and success. Their study counted access to various LMS activities and found that homework solution access is the only strong predictor of student performance. However, these studies are primarily descriptive rather than predictive.

Online teaching strategies are primarily dependent on instruction design as each mode of interaction - student/instructor, student/student, and student/content- has a positive impact on student progress. A study by Coldwell et al. [83] focused on the relationship between student participation in a fully online course and their final grades. They found a positive relationship between student participation and final grade. Dawson et al. [75] examined the impact of various LMS tools and found a highly positive correlation between discussion forum activity and student success. They observed more than 80% of interactions occurred in the discussion forum, which is the primary interaction tool in LMS. Another study by Greenland [84] found that asynchronous communication is the primary form of all online course interactions. Nandi et al. [85] found an increasing number of posts in discussion forums close to assignment and exam deadlines. They also found a high correlation between exam scores and online class participation throughout the semester, especially in high-achieving students.

All the studies discussed above adopted log files from LMS systems to extract unbiased details from activity and performance to identify a relationship between independent interaction variables and student grades. Most of the discussed studies are based on univariate analysis focusing on a single variable or a set of highly impactful variables of a single course or similar courses on student outcomes. However, student performance is a highly complex area in education to measure or understand, especially across various on-campus courses in a university setting. Most of the authors discussed above noted the need for more in-depth works to investigate student performance across courses and based on multiple variables. These studies also lack an explanation about variables used in their studies to track student performance. It is evident that the authors selected LMS variables based on their belief that these variables are highly correlated with student scores.

Factors that influence student academic performance have been the focus of researchers in LA and EDM domains for many years. It remains an active area of education research, indicating the complex problem in measuring and modeling learner processes, especially in tertiary education. Positive learning characteristics have a significant positive impact on learner engagement improvement in multiple ways. The dispositional language specifies learning as a combination of self-regulation, learning inclinations, motivation, behavioral patterns, interactions, and cognitive ability. In their study, Shum et al. [86] proposed a combination of self-reported data gathered in surveys with student interaction data generated by LMS to study individual student performance, learning processes, and group interactions. These social analytics depend primarily on student self-reported data to develop toolkits that support a specific learning type, especially in courses with high diversity [87]. However, our study focuses on objective identification of student success based on data that LMS captures. We will also identify the crucial variables from predictive model output for various student groups based on their diverse backgrounds (race, gender, and student status).

Even though there is a common agreement about the purpose of learning analytics, there are still several varying opinions on what data needs to be collected and analyzed to improve teaching and learning processes. A study by Agudo-Peregrina et al. [88] argued that it is highly complex to identify the net contribution of various interactions to the learning processes. Their findings show that peer interaction between students has a lower influence than student-teacher interaction, which contradicts earlier studies that showed high importance for student peer interactions. A survey by Dominquez et al. [89] utilized multiple variables like LMS logins, time stamps, and content access flags captured in a biology course to predict student grades at the end of course completion. The results show that the algorithm predictive accuracy is at 50% in subsequent semesters. Lerche and Keil's

[90] recent study utilized Moodle log data from 369 students enrolled in three online courses across three semesters to predict their scores at the end of the term for each course. Their regression results related to predicting student scores in a course at the end of the semester varied from 0.17 to 0.6 for all three courses. This broad range of performance across courses is due to varying content variables in each course based on the course structures. Studying the difference in instructional design, variables in extracted data, statistical inferences, predictive modeling used, interpreting model outcomes and pattern observations, etc., might explain the inconsistencies in results shown in earlier studies.

Data captured by LMS systems became prominent in LA and EDM circles as they capture student interactions in non-intrusive and ready-to-use settings. Several studies were discussed earlier in this research that utilized the LMS data to develop models that track student progress. However, it is still challenging to build highly accurate models that predict student learning outcomes across courses and understand the impact of different variables captured by LMS. Another significant gap in earlier research is their inability to predict student performance across courses in a given semester. One primary issue in predicting student performance across a semester is to find methods that aggregate student LMS variables across courses. This research shows methods to address the research gap found in earlier studies.

In this study, we approach the problem of tracking student achievement by developing student-centric models that build on aggregated LMS interaction variables collected across a semester irrespective of student year and course. One unique aspect of our work is related to the study of model performance on longitudinal student data. We develop models that predict student end-of-term GPA based on four cumulative periods in a semester. This work also focuses on explaining the impact of different aggregated LMS variables on various student groups categorized based on performance, race, gender, and student type. The importance of features is explained by adopting correlation statistics for univariate importance, a regression model for interaction effect, and LIME for model-based yet model agnostic explanations.

2.6 Chronotypes & Causal Relationships

Humans have a 24-hour internal clock running in the background that determines when to sleep and when to be productive. These 24-hour cycles are referred to as circadian rhythms. The behavioral manifestation of these circadian rhythms is termed chronotype [91][92]. Understanding an individual's chronotype helps identify their routine and provides insights about their highly active and productive time. Earlier research argues that circadian rhythms differ between individuals as a group of clock genes conditions them [93]. Nonetheless, these are not fixed and can vary during an individual's lifetime.

Students have to schedule their daily activities based on their predetermined schedules by taking social constraints like class schedules and outside work hours. General clock time preferences are more likely to match natural chronotypes when students are given more flexibility in organizing their schedules. Differences between a person's natural chronotypes and schedules influenced by social constraints cause a phenomenon referred to as social jetlag [94][95]. This social jetlag leads to an accumulation of sleep debt that causes tiredness and a decline in cognitive abilities throughout the workweek. Prior research suggested that individuals with a match in chronotypes and schedules during workweek don't show a change on weekends [96][97]. This observation is in contrast with individuals whose weekday routine varies from natural routines. As activity and sleep are both indicators of natural chronotypes, an observation of individuals on weekends might give accurate insights into their natural habits defined by circadian rhythms.

Earlier studies primarily focused on two significant chronotypes in humans: Morning and Evening. These studies showed that these two chronotypes are based on an individual's age and gender and vary across an individual lifetime. The insights from earlier work showed that children are predisposed towards morningness, but a delay of phase preference can be observed when they reach adolescence [98][99]. This shift reaches a maximum of eveningness at the age of 20. Multiple studies also observed that most individuals return to morningness again at the age of 50 [100][101]. Gender-based chronotype studies contradicted a lot in identifying variations in morning and eveningness as some studies showed women to have a greater tendency towards morningness compared to men. A meta-analysis showed that the significance between morningness and females is very weak [102]. In addition to these two demographics, other researches focus on different variables like productivity, mood, temperament, caffeine consumption, avocational interest, and internal temperatures.

Recent studies observed a new type of chronotype referred to as intermediate [103][104]. A study performed by Putilov et al. [105] in 2019 argued about four chronotypes based on time differences in alertness and sleepiness: morning, afternoon, napper, and evening. Most of these studies adopt different self-reported questionnaires that are developed to analyze an individual's daily preferences. Initially, most of these measures treated chronotypes as unidimensional: Diurnal Type Scale (DTS) and Circadian Composite scale (CCS) [106][107]. But, psychometric studies questioned this unidimensional approach to morning-eveningness [108][109]. Recent studies came up with a multidimensional approach that treated chronotypes as independent dimensions. As most of the chronotype measurement methods are self-reported surveys, biases like social desirability, recall period, sampling approach, and selective recall are hard to eliminate. Our work explores the activity patterns of students based on login times captured by the LMS to mitigate this issue. We employ clustering techniques to identify and classify students based on their similar activity patterns.

Research in psychology documented the effects of time of day on complex simple and complex cognitive functions that are based on an individual's chronotype [110][111]. An early study by Roberts and Kyllonen [112] showed that individuals who were active during the evenings did well on the measures of memory, cognitive ability, and processing speeds even though these cognitive tasks are performed during the morning time. In addition to this, this study also reported a high correlation between individuals working memory which is a proxy of general intelligence, and morningness and eveningness [111]. However, other researchers reported that the relationship between cognitive ability and chronotype is much more variegated. For example, a study by Killgore and Killgore [113] reported a significant correlation between eveningness and verbal cognitive ability but not between eveningness and math ability.

Further analysis showed that the latter finding is only observed in female participants. A similar study by Song and Stough [114] showed a significant eveningness effect on a spatial subtest of Multidimensional Aptitude Battery IQ and not on any other subtests. Most of the research in this domain is still inconsistent on which aspects of cognitive abilities do chronotypes impact.

Inspired by earlier studies that reported significant relations between cognitive ability and chronotypes, researchers in education focused on studying the relationship between academic performance and chronotypes. They mainly focused on student grade point averages and other measures extracted from their academic achievement indicators [115][116][117]. Most of the studies in this area reported that student academic achievement is strongly and inversely related to their eveningness, whereas morningness is positively associated with their achievement. These patterns are similar in students at the high school level and students at the university level. A study by Precckel and Roberts [118] showed a significant negative correlation between academic achievement and eveningness. This study was conducted on 270 German secondary students where teachers assigned grades averaged over German, Math, English, and Physics. These results seem to be consistent even after introducing controls for gender and intelligence.

A recent study focused on university students that utilized three classification scales (Morning, Evening, and Neither) found that morningness students showed much better performance on both theoretical and practical examples compared to students belonging to other chronotypes [119]. This study also showed that students belonging to neither evening nor morning chronotypes performed lower than their counterparts. Additionally, this study also observed that students who belong to eveningness did worse on practical examinations than morningness and students belonging to neither class. One interesting find in this study is related to the performance of eveningness students on theoretical exams. This study infers that the higher intelligence expressed by eveningness students compensated for their disadvantage on the theoretical exam but not on practical exams. However, it is still necessary to study the impact of an age-based shift in chronotypes with student's academic performance. This is based on the insights provided by earlier studies that reported individuals chronotype shift from morningness to eveningness by the time they reach their 20's. In addition to this, it is also necessary to study the impact of chronotypes on student-centric models that analyze comprehensive student data in a semester or academic year instead of course-specific analysis. This understanding will support the development of interventions that are effective and promise overall student success. Our work will utilize student clusters extracted from their hourly login patterns to develop and assess models that predict student performances.

The primary objective of educational research is to develop interventions that promote student academic achievement. Earlier studies in learning analytics and educational data mining focused on developing models that predict student achievements early on in their academic program to support the development of these interventions. However, most of these studies focused on course-level predictions and analysis, making it harder to develop interventions at the student level. Another gap observed in prior research is the lack of causal understanding related to features that causes a shift in student academic performance. Causal inferences promote the development of impactful intervention techniques that positively impact student academic performance.

Traditional research utilizes randomized control trials to discover causal knowledge. However, these are time-consuming, expensive, and sometimes unethical and impractical to implement with real-world students. Studying causal relationships from student interaction with LMS systems allows us to explore a vast amount of noninvasive data collected by these systems. This method will provide a search for answers beyond common correlations. Yet, this type of analysis is not common in learning analytics and educational data mining domains. For example, a study by Fanscali [120] investigated the role of carelessness in analyzing the counterintuitive relationship between students' learning outcomes and their affective states related to confusion and boredom. This study utilized observational data from the algebra cognitive tutor platform. An earlier study by Fanscali [121] on a similar dataset showed a relationship between student's affective states, gaming the system, prior knowledge, and their learning outcomes. This study utilized a causal framework termed causal discovery with the model. In their research, Koedinger et al. [122] analyzed interaction data captured by the online learning environment. This study reported causal interaction between active student engagement and their learning outcomes. Inspired by this work, a recent study by Chen et al. [123] developed a causal discovery framework that utilized TETRAD [124], a causal discovery and inference toolkit. Our work adopts the framework proposed by Chen et al. [123] to study causal relationships between student login behaviors captured by LMS and learning outcomes.

Chapter 3: Affective Dimensions of Learning based on Digital Learning Environments

Understanding student learning behaviors is of the prime importance of educational research. However, many complex factors influence learning processes, but one collective impact of all these factors is on human affect that influences learning and degree of motivation. This chapter discusses the current state of human affect detection in education, our proposed affect change model, and its implications. This study adopts a dataset from ASSISTments online learning platform consisting of student interaction data, and ground truth labels for affect states coded by BROMP certified coders to develop and validate the affect change model. We show that the proposed affect change model combined with the adoption of machine learning algorithms will support the development of a ubiquitous learning system that tracks the student learning process in context of contributing factors and provides interventions when needed.

3.1 Introduction

The increase in the use of intelligent educational technologies such as learner management systems that collects a vast amount of student interaction data enabled researchers to develop new applications that empower student learning abilities. The development of a sensing system that tracks student learning poses unique opportunities and challenges. One such challenge is developing an automated system that tracks human affect that impacts learning during their system interactions. The affective process has a substantial influence on the learning processes in students, including attention, significantly modulating the selectivity of attention as well as motivating learning action and behavior.

BROMP has explored these affective processes and identified four fundamental affect states related to attention; concentration, confusion, boredom, and frustration [125]. Based on the protocol, computational models for detecting these affect states based on student interactions with computer systems became prominent in recent years. Studies focusing on the development of affect state detectors based on human-computer interaction and physiological sensors showed some promising results [71][126][127]. However, studies in psychology argue about relevant emotional theories that can assist in the prediction of student affect states from their system interaction [128][129]. Therefore, this work revisits previous computational efforts and datasets and then explores novel computational models focusing on the characteristics of affect dynamics.

Affect dynamics and affect chronometry are the two areas that focus on temporal changes in affect states. Affective dynamics support the understanding of frequent affect state transitions over time. D'Mello and Graesser modeled the impact of these temporal changes in students during complex learning tasks [130]. This study shows that a student from engaged concentration state transitions to confusion by experiencing cognitive disequilibrium when facing anomalies, contradictions, obstacles, and any other impasses. Restoration of equilibrium helps them get back to concentration and avoids their transition into the frustration that changes to boredom with time.

Affective chronometry is a part of affect dynamics that focuses on individual affect transition over time [131]. This mainly deals with response latency that focuses on the time it takes for an emotion to reach its peak from the onset and recovery time is when the emotion dissipates [130]. From the study on the half-life of cognitive affect states that focuses on the recovery time of emotions of students, it is evident that the decay rates for affect states like concentration, boredom, and confusion were lower compared to frustration and delight [132]. The increasing interest in understanding affect dynamics and affect chronometry from Quantitative field observations like BROMP and sensory data poses new sets of challenges and limitations [125].

The increased interest in leveraging affective computing to assess the students' affective states has generated many successful approaches to building accurate affect detectors. Specifically, along with the broad spreading of ubiquitous computing systems such as mobile phones and wearables, the physical and physiological sensors provide a window into the relationship between everyday behavior and affective states [133][134]. Although these sensors are getting popular, the affect detectors that rely on these sensors still have challenges in scalability, especially for those students who cannot afford mobile phones and wearables. The situation might cause limited education benefits for the students from low-income families. Therefore, affect detectors that solely rely on the learning data captured by the computer-based learning systems may be more feasible to deploy in scale to provide equal educational benefits for those students in the classroom.

Sensor-free affect detectors have been investigated for many years, and the models learned from previous studies have been verified and validated in the diverse dataset [72][135]. Currently, the best sensor-free affect detector has demonstrated better performance than a chance to detect student engaged concentration, confusion, frustration, and boredom solely from students' log data. The datasets used in these studies include the cognitive tutor dataset and the ASSISTments dataset [72]. The detectors built from these datasets helped various researchers to study the influences of students' affective states on multiple aspects. Meanwhile, much attention has been put on increasing the accuracy of sensor-free affect detectors and eventually using them to drive intervention, such as data cleaning for missing skills and wrong answers [72], new training models leveraging the deep learning techniques [71]. However, little has been reported to develop affect detectors leveraging existing psychological theory [132][136][137]. For instance, researchers in affect styles have demonstrated that the temporal features of affect (e.g., affect change) include richer information than affective states and expose the individual progress in affective regulation.

3.2 Our Work

In this study, we proposed an affect-change detection model built based on the concepts involved in affect dynamics and chronometry [138]. This work seeks to apply these existing psychological discoveries to develop better affect detectors solely based on the learning data captured by the learning system. Previous psychological studies developed models to study the temporal features of the transitions between two types of affective states; positive and negative [136]. Our work adjusts the model to fit the four types of affective states; engaged concentration, frustration, confusion, and boredom, including six types of transitions.

In this study, we explore the application of this model to detect affective states using data collected in the context of the ASSISTments online learning platform. After statistical analysis of the dataset, we focus on four types of transitions, (concentration - concentration), (concentration - frustration), (concentration - boredom), and (concentration- confusion), which is equivalent to previous affect detectors [72].

Deep learning models are still being investigated in the education domain for prediction of student affect states but were extensively adopted to extract student knowledge states [71]. Inspired by these capabilities of deep learning to extract features from data and provide efficient classification, we adopt deep Convolution Neural Networks (CNN) and variations of Recurrent Neural Networks such as Long short-term memory (LSTM) and Gradient Recurrent Unit (GRU) to classify student affect state changes. In this work, we train and test conventional machine learning models and deep learning models to detect student affect state transition based on their interactors with the online learning platform.

3.3 ASSISTment Dataset Analysis

In this study, we adopt a dataset consisting of student interaction data with ASSISTment online learning platform and affect states coded by BROMP certified coders. ASSISTments is an online math tutoring platform developed and maintained by Heffernan and a group of graduate and undergraduate students at Worcester Polytechnic Institute. Ken Koedinger from Carnegie Mellon University is a collaborator on this project and is credited with this development. This platform helps students increase their learning ability and gives immediate feedback employing interventions so that teachers can assess the progress of students [71][139]. MIT highly recognizes this platform as one of the best computer-assisted intervention platforms that help students in learning math based on reviewing more than 100 institutions [140].

3.3.1 BROMP for Data Collection

BROMP Protocol was first used in 2004 by Baker to record appropriate behaviors in an educational setting, and later affective states were added by Rodrigo et al. to this protocol in 2007 [141][142]. Trained and certified human coders utilized BROMP to collect the affect states used in this study [125]. They capture the affect states by observing students during their engagement with the ASSISTments platform. The student affect states are respectively coded into five categories in the ways of a round-robin fashion for every 20-second interval for each student comprised of confusion, engaged concentration, boredom, frustration, and impossible/other to code and cover the whole class [71][125][126]. For every observation, it is necessary to record time stamps and identify missing labels. Temporal features are considered while reorganizing this data, which will be discussed more in the following sections of this study [138]. For an in-depth understanding of BROMP coding, readers can refer to the BROMP training manual [125].

The dataset used in the research consists of 756 students from different demographic regions like rural, suburban, and urban. There is a total of 7663 successful field observations so that each student is observed at least ten times on an average [71][125]. A total of 204 feature vectors are collected after cleaning the data, including the response labels. These features mostly contain system usage time, usage of hints, scaffolding, and many other features to capture student's low-level interactions with the learning platform [138].

3.3.2 Data Imbalance

Previous studies on this dataset have mentioned some concerns with the imbalance issues in this dataset and proposed useful resampling methods to solve these issues [71][72]. These resampling methods are not detailed, and most of them did not explicitly discuss the imbalance issue. This non-uniformity of the data set will lead to class imbalance issues in machine learning and further biases in the classification process. Statistical observations from the dataset show that approximately 80% of the states identified in the data set are engaged concentration and the remaining 20% of the dataset consists of boredom (12%), confusion (4%), and frustration (4%). One primary reason for more concentration labels in the dataset is justified by earlier studies that discussed student interest in working with the ASSISTments platform [71]. One other reason can be the faster decay of affect states such as frustration and confusion compared to concentration and boredom that might not be captured by BROMP coders [136].

Due to the persistent imbalance issues in this dataset, the machine learning models might overfit during the training with more bias towards higher sampled classes like engaged concentration [143]. Resampling methods are developed by previous studies using up or down sampling the minorities and majorities in data, assuming that the hidden features are extracted from data with interpolation techniques in these methods. These techniques are hard to verify during model training and testing because the effects of traditional interpolation techniques are difficult to understand or validate [144][145].

This prospective study conducts a statistical analysis to identify the transitions among these four affective states, as shown in Table 3.1. 70% of students working on this platform always concentrated without transition to any other state during field observations. There are no students that are always frustrated in the dataset, and this can be related to an earlier study that shows frustration as a transient state with a faster decay. Less than 1% of students have transitions between confusion and frustration. The major transition of states is between concentration to boredom at 17%, followed by concentration to confusion and concentration to frustration, as shown in Figure 3.1. The transition between minor states boredom to confusion and confusion to frustration accounts for less than 3 % of the total transitions in the dataset. It is coded as weaker transitions, as shown in Figure 3.2.

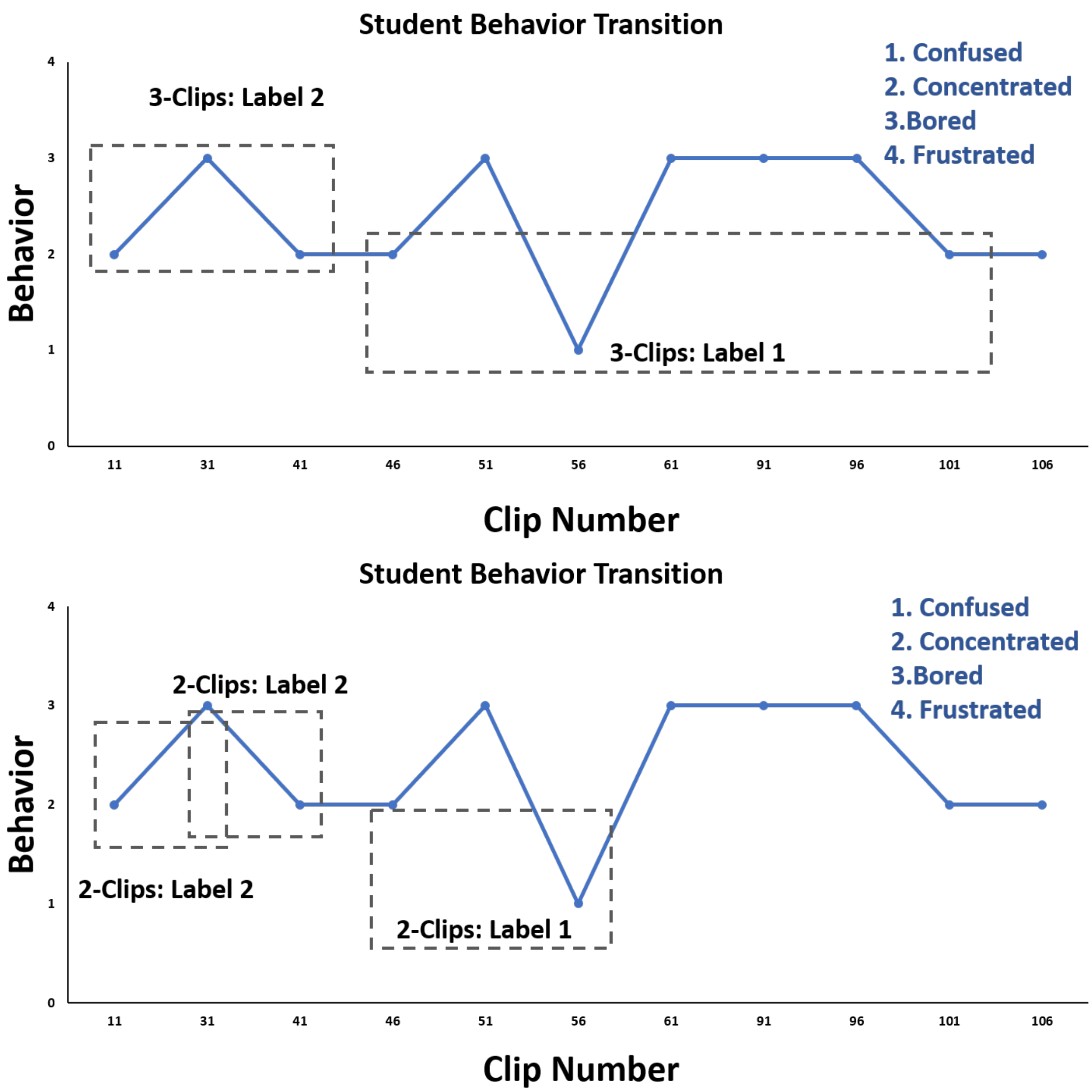


Figure 3.1: Affect transitions of a student during learning over time

From the statistical analysis of this dataset, we find two implications, and the first one is to simplify the models into four major transitions (Always Concentrated), (Concentration <> Confusion), (Concentration <> Bored), (Concentration <> Frustration).

The second one is to downsample the students who are always concentrated to solve class imbalance issue that reduces the influence of clips without transitions in affect for a reliable feature distribution [138][146].

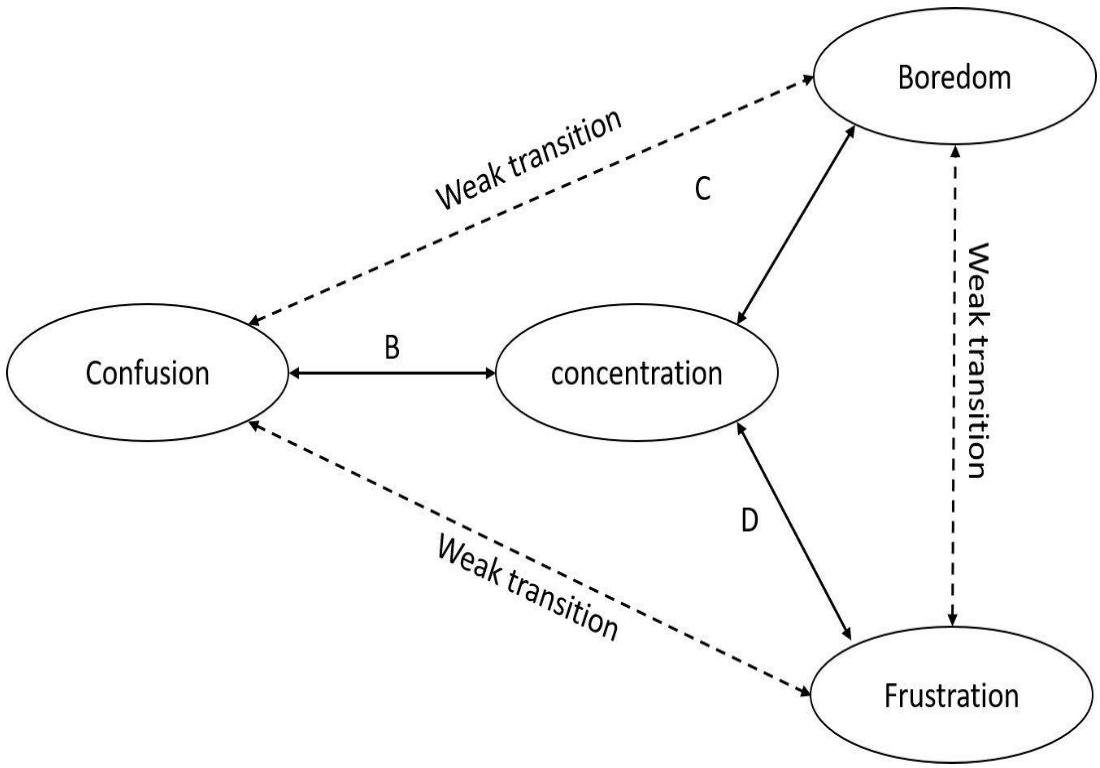


Figure 3.2: The affect state change model with the transition between different affect states

Table 3.1: The students who did/not experienced affect changes

Affect Change	Number of students	Percentage of students
Always Concentrated	511	67.60 %
Always Bored	7	0.93%
Always Confusion	1	0.13%
Always Frustration	0	0%

Concentration <> Confusion	58	7.67 %
Concentration <> Bored	126	16.66 %
Concentration <> Frustration	53	7.01 %
Bored <> Confusion	15	1.98%
Confusion <> Frustration	7	0.93%

3.4 Methodology

In this study, the first step is to reorganize the labels into 2 and 3 clip structures. A correlation-based feature selection algorithm is applied to extract essential features that support predicting affect state transitions [147]. These extracted features were then used to train and test six conventional machine learning algorithms (Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Neural Nets (NN), SVM and AutoMLP (AMLP)) and four deep learning algorithms like (Recurrent Neural Network, LSTM, GRU, and CNN). For training these algorithms, we adopt RapidMiner for conventional machine learning algorithms and keras for deep learning algorithms.

3.4.1 Reorganize and Relabel the Dataset

The proposed approach reorganizes and relabels data samples based on the affect change model mentioned in the above sections. The 2-clip and 3-clip datasets are organized based on the affect dynamics concept discussed in the above sections are shown in Figure 3.3. This student has 11 samples/clips labeled by BROMP coders. These samples were then reorganized into 3 clip and 2 clip datasets, as shown in Figure 3.3. We observed a single transition in data based on the affect chronometry concept that focuses on transition in affect state. We reorganized labels in the form of 2-clip data samples that consists of a subject affect transition from one state to the other. The 3-clip data organization is based on affect dynamics that focuses on the affect cycle, which considers concentration as the initial phase and then transitions to another affect like confusion and ends at the concentration.

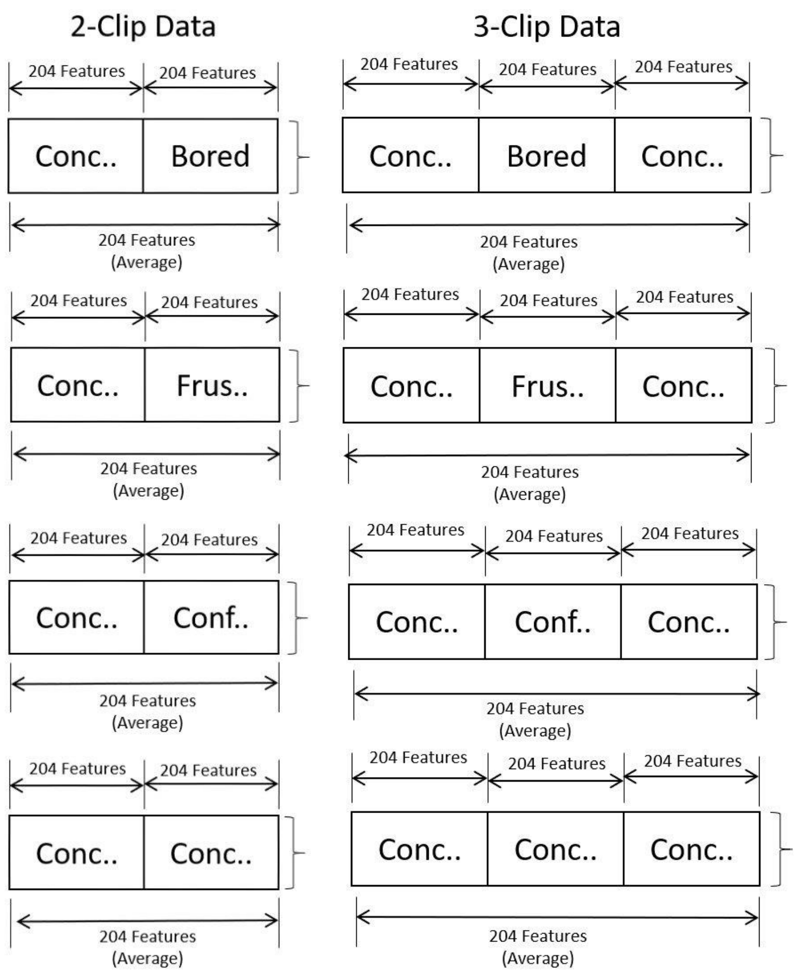


Figure 3.3: Illustration of the organizing and labeling process of 2-clip (Left) and 3-clip (Right) datasets.

Once the datasets are reorganized based on the affect change model, the 2-clip dataset consists of 408 features (204*2), and the 3-clip dataset will have 612 features

(204*3). The dimensions of these datasets were reduced based on an average function that averages similar columns in 2- clip and 3- clip data that reduces the final dimensions to 204 features. The primary reason for not utilizing dimensionality reduction techniques like PCA or ICA is the difficulty in understanding the role of each feature on the predictions as principal components are formed based on all the features in data [148].

3.4.2 Feature Selection and Model Training

This study adopts a feature selection technique based on correlation, ID-ness, and stability [147]. Pearson correlation between attribute and target label is calculated. ID-ness and stability are reciprocals that define the unique and constant nature of a given attribute. An attribute with higher ID-ness indicates the column values are distinct and lower ID-ness indicates similar values. Feature selection takes place based on a set cutoff for these values, a feature with a correlation higher than 0.01 percent, ID-ness less than 85%, and stability less than 90% are selected. The dataset adopted in this study is free of missing values, so this measure is ignored in feature selection.

This study adopts the RapidMiner data science platform to train and test conventional machine learning algorithms and keras in python for deep learning algorithms [149]. A five-fold cross-validation method is used to split data at the student level to train and test all samples present in data and extract reliable performance metrics that can be generalized over the data. This approach creates a binomial classification problem by dividing the affect change model dataset into four datasets with a major class (always concentrated) and a major to minor class transition.

Training a neural network in smaller batches and updating the weights of a network at the end of each training cycle is a common practice. The model is updated by Adam optimizer at the end of each training cycle. The affect change model mainly focuses on the time series nature of data. Still, the temporal nature of data is inconsistent as the BROMP coders are not continuously labeling a student. CNN focuses on piece-by-piece information of an image that it tries to learn from the training data and compare with the testing data. These networks are least concerned with the time-series nature of the data compared to RNN, whose dependency is mainly on the temporal characteristics of the dataset [150]. The inconsistency in this data might hinder the performance of Recurrent neural networks.

Model training in neural networks happens in a multitude of epochs where the network is training multiple times to extract as much information as possible. This multiple training also leads to the overfitting of a network that performs poorly on unseen data. To reduce this issue, we adopt a variable epoch method that focuses on the performance on hold out data after each training epoch to see if there is any improvement in the model performance. The network stops training when it sees no improvement in the performance on hold out dataset. Average cross-entropy is the selected performance metric calculated after 15 epochs to see if there is an improvement in the training performance. The model training stops when a higher value of this metric is found compared to the previous value. Sixty epochs seem to be the best number over all the models.

3.5 Results

This study considers three performance metrics, Area under the ROC curve (AUC), Cohen's kappa, and root mean square error (RMSE), to evaluate the developed models. This study considers a trade-off between all the three performance metrics to decide on a better algorithm. The algorithm that performed better in detecting the 2-clip affect changes is shown in Figure 3.4. Support vector machines outperformed other algorithms in detecting affect state changes from concentration to boredom vs. all other transitions and always concentration vs. all other transitions. Affect transition from concentration to frustration and confusion is detected very well by single layered feedforward neural networks with kappa values of 0.296 and 0.31. SimpleRNN performed better compared to LSTM in detecting all transitions except concentration to boredom.

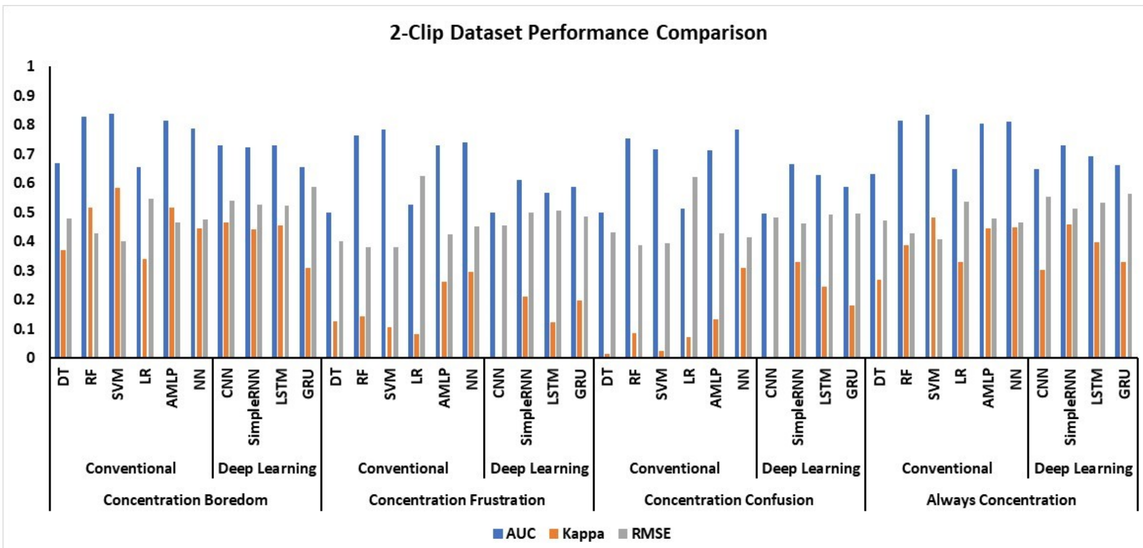


Figure 3.4: Prediction performance of different statistical models on 2-Clip affect change dataset

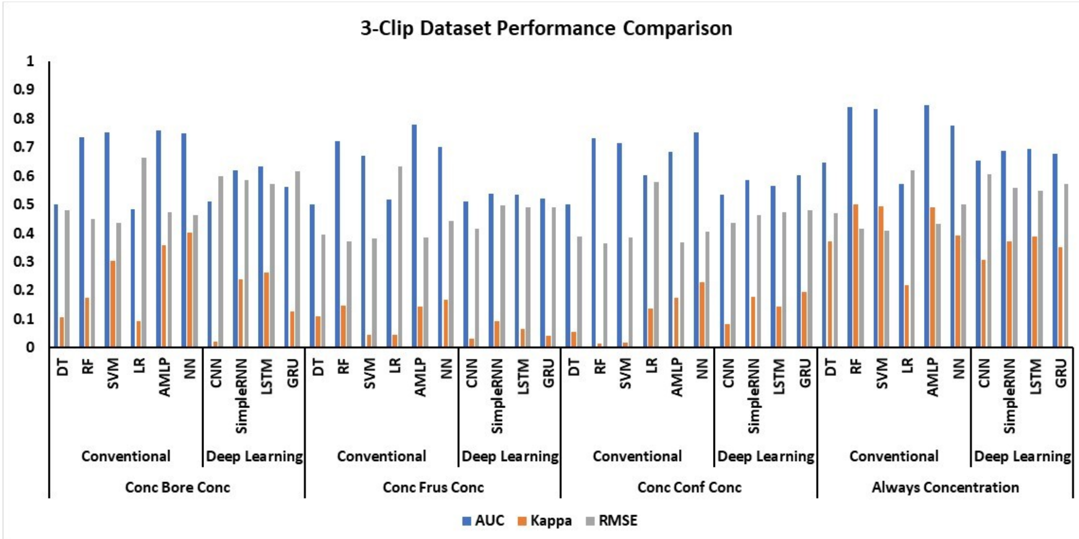


Figure 3.5: Prediction performance of different statistical models on 3-Clip affect change dataset

The 3-clip affect transition data performance results are shown in Figure 3.5. Neural networks and AutoMLP performed better for datasets with frustration and Boredom, whereas random forest performed better in detecting always concentration. Tables related to Figure 3.4 and Figure 3.5 are available in Table A. 1, Table A. 2, Table A. 3, Table A. 4, Table B. 1, Table B. 2, Table B. 3 and Table B. 4 listed under Appendix A and Appendix B.

3.6 Discussion

Although affect change has been a well-studied topic in psychological studies, this domain knowledge has not been given enough attention in education [151][152]. Although affective states have attracted much attention recently in developing an intelligent tutoring system, the bridge between the domain knowledge and the objective data captured by the learning system has not been established well [71]. This work attempts to develop an affect-change model to build the relationship between the domain knowledge and the learning dataset previously studied using traditional feature engineering and machine learning algorithms.

3.6.1 Findings

From the study, we find that the performance of detectors with conventional machine learning algorithms is better than the deep learning algorithms. The AUC values are higher for conventional machine learning algorithms when detecting affect state changes. The kappa values are higher for traditional algorithms except for state change detection of concentration to frustration, where the kappa values are higher for deep learning during 2-clip analyses. The two algorithms perform very well during the 2-clip analysis of state change from concentration to boredom. The best models of conventional machine learning outperform deep learning to affect state change detection with high AUC and kappa values.

Most detectors are accurately concerned with changes from data that is an average of a feature of a different state. We only present the best algorithm that acquiring the highest AUC and kappa in 6 conventional algorithms trained and tested by us. It also concludes that algorithms in deep learning perform well with different changes in the selection of parameters. This trade-off should be considered during the selection of a model for detecting different affect state changes. From the results, we can see an improved performance in the deep learning model during 2-clip affect change detection from concentration to boredom. Besides, we also observe that the confidence levels of every algorithm are higher than the threshold of 50% in the research.

In comparing the two approaches, the key advantages of the conventional machine learning algorithms are high interpretability and their performance is better than deep learning models. However, the deep learning algorithms with higher kappa values perform better in contrast to previous studies of affect detection [71][72][126]. Deep learning algorithms perform better with vast amounts of data that is hard to collect with human observation coding like BROMP. To facilitate the advantage of dealing well with time-series data, deep learning could be employed to analyze the data acquired from sensors which can be enormous. One of the disadvantages of the deep learning algorithm is that they are highly sophisticated when compared to conventional algorithms, and it is difficult to analyze and interpret the model parameters

The performance of the affect change model was better than previous work in affect state detection, as shown in Table 3.2 [71][72][126]. The comparisons in Table 3.2 are made between the best detectors in each model with high AUC and Kappa values. Meanwhile, the performance of the detectors is more balanced than previous methods. For instance, the performance of each detector in state change is better than all the previous three studies [71][72][126] and detectors based on deep learning demonstrate unexpected performance such as AUC is higher as 0.72, but kappa is 0.09 [71]. Overall, our affect-change detectors perform consistently to all of the affect changes.

Table 3.2: Comparison of Affect State Vs. Affect Change model performance

Affect State Vs Change	**Model**	**AUC**	**Kappa**	**Study**
Engaged Concentration	SVM	0.837	0.479	Our Study (State Change)
	Feature	0.743	0.423	Wang et al., 2015
	LSTM	0.80	0.34	Botelho et al., 2017
	Logistic Regression	0.624	0.139	Jiang et al., 2018
Confusion	Neural Net	0.783	0.310	Our Study (State Change)
	Feature	0.625	0.148	Wang et al., 2015
	LSTM	0.72	0.09	Botelho et al., 2017
	Logistic Regression	0.568	0.091	Jiang et al., 2018
Boredom	SVM	0.838	0.585	Our Study (State Change)
	Feature	0.671	0.260	Wang et al., 2015
	LSTM	0.80	0.28	Botelho et al., 2017
	Logistic Regression	0.682	0.278	Jiang et al., 2018
Frustration	Neural Net	0.739	0.296	Our Study (State Change)

Feature	0.602	0.157	Wang et al., 2015
LSTM	0.76	0.15	Botelho et al., 2017
Logistic Regression	0.634	0.056	Jiang et al., 2018

Furthermore, our models such as neural nets and SVM have fewer parameters than deep learning methods but achieve reliable performance in detecting all affect changes. The detectors trained by the reorganized dataset obtain better AUC and kappa than past results reported on the condition of the same dataset. The improved performance is consistent with our hypothesis that the affect changes contain richer information than single affect states. Furthermore, our detectors perform much reliable in detecting all the affect changes with higher AUC and higher kappa. This reliable performance points out that the affect-change detector may be the solution for research and practice in sensor-free affect detection, eventually guiding to realize affective tutoring system applied to drive affect-sensitive intervention.

3.6.2 Challenges

From the findings mentioned above, we also encounter multiple challenges in this study.

(i) Scalability: The current system uses BROMP methodology that involved a human coder to code student affect states that reduce scalability.

(ii) Missing Data: As the coding of student affect states happens in a round-robin fashion, there is a lot of unlabeled temporal data.

(iii) Integration: Current systems in the domain don't integrate data well as there is no direct link between the human coder and the learning management system.

To address these challenges, we consider working on a ubiquitous sensing and learning system that integrates learner management systems and predictive modeling to track and improve student learning abilities.

These emotions can be captured at any time without human coding and also reduces missing data. With the availability of high-speed internet, data from the sensors and learner management systems can be integrated efficiently into the cloud platform. These cloud platforms are also compatible to run complex algorithms on the data as they provide scalable resources.

Our proposed affect change model deals with transitions in human affect, and these transitions will reflect individuals' physiological measures [133]. Current mobile and wearable technologies have the capabilities to detect transitions and boundaries that can be segmented in real-time [153]. There are also change point detection methods that support detecting changes in physiological measures and are successfully implemented in activity recognition systems [153]. The detection of changes in affect helps build a wearable-based system that can notify or deliver interventions to users with fewer interruptions and reduced cognitive load.

This chapter has discussed a research study involving students in middle school mathematics who experienced a web-based system that aims to provide immediate feedback to the many students in the classroom and at home every day [71]. The proposed affect change model proves to be efficient and helps support the development of intervention strategies that improve student learning. This model also leads to developing a Ubiquitous learning system that tracks student affects with minimal human intervention and mitigates the challenges in the current system.

Chapter 4: Student Outcome Detection & Factor Analysis from Digital Learning Environments

Exploring predictors concerning whether a student will enroll in a STEM major has been investigated for understanding and further facilitating the processes that lead students to become interested in and equip them for STEM careers. Existing research focuses on isolated aspects of the student's interactions with the online learning system, such as student knowledge and affect states. However, these factors could be optimally selected and combined to improve the prediction performance of student career choice models as they influence one another. This chapter proposes a voting-machine approach to predict individual student career selection based on knowledge, affect states, and clickstream data recorded during their interaction with the ASSISTment learning platform. The features were evaluated and selected based on correlation, ID-ness, Stability, and Missing Values. The selected list consisting of high and medium quality features was then used to train and test different machine learning algorithms. The trained models will follow the voting-machine mechanism to determine the predicted l-bels and their confidence. The results showed that Gradient Boosted Tree and KNN predict student career choice on both training and testing datasets with better performance compared to existing models. This proposed method helps in the early identification of student career selection and supports the development of intervention strategies that encourage them to choose STEM fields.

4.1 Introduction

Understanding the processes that lead students to become interested in and equip them for STEM careers is a critical step in developing better programs to prepare students to enroll in STEM programs. Therefore, with recent advances in the development of online learning platforms like MOOC and intelligent tutoring systems (ITS) that enable researchers in education to study various factors that influence the prediction of career selection [70][154], multiple models were developed in this domain to understand the reasons for whether or not a student is interested in STEM fields [155]. These digital learning systems are designed to capture data regarding the students' interaction with the system or software, such as click-stream records and learning outcomes. Then through the development of the experimental protocol and theoretical models, the data patterns could be classified and identified to link with meanings within educational theories, such as self-regulated learning behavior and the affective processes. One significant finding from past research is the capability of student knowledge states in mathematics during their middle school years have a considerable influence on their choice of future career [69][156][157][158]. However, this finding does not consider much about the affective process of the learning behavior, which has been proved as a critical competency of student learning.

Researchers argued that the integrative perspective regarding the student knowledge and their affect states inferred from the clickstream records [159] should empower the performance in predicting whether a student will enroll in a STEM course. Specifically, previous studies show evidence that student emotional state will influence their behavior state, impacting cognition while learning [64][160]. These affect states of the student will also control their interaction with ITS and MOOC systems [161]. Regarding the interplay between academic emotion and student engagement, Pekrun & Linnenbrink-Garcia suggest that student engagement impacts the relationship between learning and emotions [160][162]. These studies also show different emotions can act as either inhibitors or propellers for student engagement in learning [64][160][162]. These changes in students' engagement during their learning process will hinder their progress in learning and impact their decision-making.

Previous studies worked on predicting a student's career choice by extracting their knowledge state from multiple ITS data samples and incorporating all the information into a structured feature integrating knowledge states [70][156]. These features only consider the average student affect states rather than individual affect states data [70]. In contrast, our work considers every interaction of the student with the platform. It uses them as samples for prediction, which enables us to capture rich information available in all the samples, including knowledge states, affect states, and other clickstream features. These samples are then used in the prediction and evaluation of different machine learning models.

This chapter proposes a voting-machine-based mechanism to predict individual student choice through optimal feature selection that incorporates features from knowledge states, affect states, and other clickstream features. The feature selection technique considers Correlation, ID-ness, Stability, and Missing Values of every at-tribute [149][163][164]. These features are then used for training and testing different machine learning algorithms. All the predicted labels for each student were used to develop a voting machine mechanism to calculate individual student STEM and Non-STEM prediction confidence. This confidence was used to extract various performance metrics of algorithms.

The highly supporting and contradicting features for each sample dataset were also identified based on neighboring attribute weights that utilize correlation as an identification factor. The local linear relation between attributes is highly influential in predicting the non-linear global relationship [165][166]. Analysis of features supporting and contradicting predictions related to STEM, Non-STEM classes facilitates understanding the importance of each feature on individual class prediction.

4.2 Data

ASSISTments is a web-based mathematics tutoring system designed for middle school students and actively maintained by Worcester Polytechnic Institute. This software records all student interaction that is useful in modeling their knowledge and detecting their engagement. Bayesian Knowledge Tracing (BKT) algorithm was applied to the data to model each student's knowledge. This algorithm provides students knowledge based on their previous responses to questions related to the same skill set. BKT assesses the knowledge state of a student at every attempt they made [167]. Affect states of students and their disengaged behaviors such as concentration, boredom, confusion, frustration, off task, and gaming was estimated by proven detectors [69][156] using the interaction data captured by the ASSISTments platform. The design of these affect state detector changes for schools located in urban and rural areas [156].

Student behavioral data captured from ASSISTments online learning platform during academic years 2004-2005 to 2006-2007 was made available during Educational Data Mining (EDM) Competition in 2017. This data consists of knowledge, affect state, and system interaction information related to 1709 students [156]. In this study, only 591 students were considered as they were labeled as STEM or Non-STEM. From the analysis, we observe that there are 316974 samples of interaction data related to these 591 students, of which 125 students enrolled in STEM careers and the remaining 466 students enrolled in Non-STEM careers. The distribution of samples in this dataset is shown below in Table 4.1. We did not apply any downsampling methods on this dataset as these types of class imbalances need to be dealt with in the real world.

Student profile data consists of System usage, Number of actions, affect states, disengaged behaviors, and abilities [70].

- Affect states & Disengaged behavior: Clickstream records were used to derive four affect states (Concentration, Confusion, Boredom, and Frustration) [127]. We utilize the rescaled value of affect states provided in each student sample. We did not incorporate the average affect state data as we work with individual samples of students.

Table 4.1: ASSISTments data distribution statistics

Career Enrolled	Students	Samples	Distribution (%)	Average
STEM	125	64719	20.4	518
Non-STEM	466	252255	79.6	541

4.3 Methodology

We divided this study into four phases. The first phase deals with feature selection. These features are optimally selected by incorporating domain knowledge and multiple statistical parameters. Once relevant features were chosen, we then used the EDM Competition's dataset to train and test various machine learning algorithms with 5-fold cross-validation. We then apply the model from cross-validation on the testing set provided by the EDM Competition. We then evaluate the performance of these algorithms and choose the best one for explaining predictions. As the cross-validation trains and predicts the output for every sample in the dataset, these predictions were used to calculate the confidence for individual student prediction. Figure 4.1 below shows each step in the analysis. The ASSISTments platform collects this data during student interaction. Student interaction data is utilized by the feature selection technique to select appropriate features for prediction, which are fed into predictive models. The proposed voting machine mechanism uses the model predictions to calculate the confidence of each class and provide a predicted label.

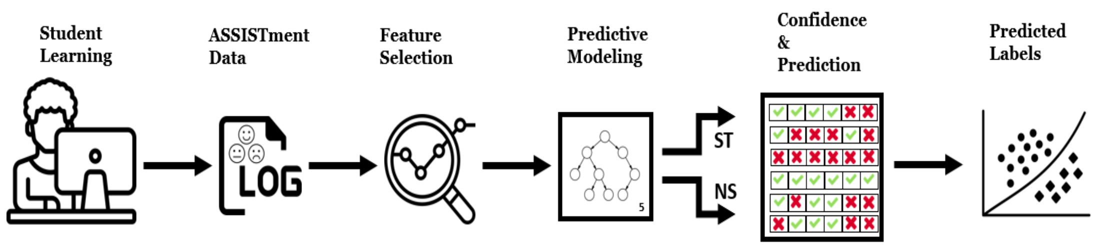


Figure 4.1: The above process pipeline represents various steps in predicting individual student career choice. Icons by Freepik, Baianat, Trevor Dsouza, Noura Mbarki & Dave Gandy from www.flaticon.com. Licensed by Creative Commons BY 3.0

4.3.1 Feature Selection

This study applies feature selection based on three significant measures related to data in the attribute. We consider attribute correlation, ID-ness, Stability, and Missing values, but the measure of the missing value was ignored as the data consists of very few missing values [147][165]. We also removed some attributes based on their similar characteristics with other columns. For example, in the dataset, we observed two confidences for affect states original and rescaled. From this, we select the attributes related to rescaled confidence. The selected attributes are also labeled based on their impact. We apply this feature selection technique adopted from the RapidMiner data science platform [149] to choose a subset of features that are useful for prediction.

Correlation is the amount of similarity between the current attribute and the target column that needs to be predicted. ID-ness is the uniqueness of values present in the attribute. For example, it merely suggests if the column is an identity column based on the values. This ID-ness is only applied to columns that have positive integers. The stability measure indicates the percentage of identical values present in the column. As discussed above, we removed missing values statistics from the analysis as we observed that there are no missing values present in the selected attributes.

The attributes are divided into three classes high (H), medium (M), and low impact, based on the measures mentioned above. These attributes are shown in Table C. 1, Table C. 2 and Table C. 3 listed under Appendix C. The high impact attributes are the ones with medium correlation, low id-ness, low stability, and no missing values [147][164]. The attributes that are classified as medium impact are based on the correlation. Attributes with less than 0.01% and greater than 50% correlation are set in this category. The low impact attributes are categorized based on stability, missing values, and ID-ness. Attributes with more than 70% missing values, ID column, and more than 90% stability are classified in this category. In this study, we removed attributes that are classified into low impact as these are trivial in prediction. We considered 40 attributes that belong to medium and high impact categories for predictions.

4.3.2 Model Training and Testing

This study trained and tested different machine learning algorithms to compare their performances on the RapidMiner data science platform [163]. This platform was recognized as a leader in Gartner Magic Quadrant for data science and machine learning for six straight years [168]. We adopt a 5-fold cross-validation method in which every sample was used for both training and testing. We selected seven predictive algorithms for this study, of which four are related to traditional machine learning, and three corresponds to neural networks. Gradient Boosted Tree (GBT), K-Nearest Neighbor (K-NN), Decision Tree (DT), Random Forest (RF), Neural Net (NN), Deep Learning (DL), AutoMLP, and Logistic Regression (LR) were evaluated on this select-ed feature. We used default configurations for all algorithms, but for GBT, we increased maximal tree depth to 20, and for deep learning, we applied two fully connected layers with 256 neurons each. These are the best performing configurations from our analysis. We compare the model performances based on AUC, Accuracy, Kappa, Root Mean Square Error (RMSE), and linear aggregation of AUC and RMSE values.

4.3.2.1 Models & Hyperparameters

- Gradient Boosted Tree: A gradient boosted tree algorithm is a sequential learning algorithm in which a subsequent tree learns from the weak predictors of a previously built tree. The tree adopted in this study has a maximum of 20 trees, a maximal tree depth of 20, and a learning rate of 0.1.
- Random Forest: A random forest is an algorithm that works based on the ensemble learning principle. This algorithm can combine different models developed based

on the bagging method. We obtained optimal settings for this algorithm with a maximum of 100 trees and a maximum depth of 10 per tree.

- Logistic Regression: Logistic regression method is for classification problems as it predicts the probability of each class and classifies based on the probability values. We adopt the standard settings for this model.
- AutoMLP: A multilayer perceptron is a feed-forward neural network that consists of multiple hidden layers in training a neural network. The AutoMLP algorithm can set the optimal learning rate and hidden layers during training. This algorithm works on stochastic optimization and genetic algorithms. This algorithm trains small ensemble methods in parallel with different hyperparameter settings like hidden units and learning rate, which are validated to find the best setting.
- Deep Neural Network: A deep neural network is an algorithm that can work with different activation layers, learning rates, and optimizers. This study adopts a four-layer (input, hidden_1, hidden_2, and output) fully connected deep learning network with 250 hidden units in each layer. We set the learning rate at 1.0E-5 and use the rectifier activation function. The regularization parameters were auto-adjusted based on the training performance of the algorithm.

4.3.3 Voting Machine based Classification

One issue dealing with multiple samples per student is to converge to a single predicted label. In this study, we propose a voting machine-based mechanism [169] that converges to a single predicted label per student based on the confidence of each class. To calculate the confidence of each class per student, we count the number of samples that were predicted as STEM and Non-STEM. The confidence value per class was then calculated by taking the ratio of predicted STEM labels to total samples per student and predicted Non-STEM labels to total samples per student. The predicted label is then assigned based on the class with the highest confidence value. Once the predicted labels were assigned to all students based on the mechanism mentioned above, the performance metrics of each algorithm were generated by comparing the predicted label and the actual label of all students. This mechanism is followed to calculate performance metrics for both training and testing datasets.

4.3.4 Predictor Explanation

The purpose of this study is to understand the factors that influence STEM and Non-STEM predictions. For this purpose, we utilize a novel prediction explanation method based on the Locally Interpretable Model explanation (LIME) [170] to find attributes that support and contradict predictions. This method creates neighboring data points for each sample in a dataset and calculates local correlation values to identify the weights of each attribute. The predictors that support and contradict are classified based on the local correlations and weights calculated for each attribute for every sample. The words "supporting" and "contradicting" refer only to the predicted value, which might be an accurate or inaccurate prediction. The linear relationship between attribute and prediction locally is highly influential; even the attributes are nonlinear globally. This method is developed in conjunction with RapidMiner software and discussed below.

LIME explains each prediction made by a complex model by training a surrogate model locally. In this way, LIME tests the predictions based on variations in data. This process is done by generating new data samples around the observation by permutations. The complex model that needs to be understood is used to make predictions. Once the predictions are made, LIME trains the linear surrogate model and weights the samples of interest. The model explains the predictions locally, which might not be good globally.

$$Explanation\ (x) = \arg \underset{g\ \in\ G}{min}\ L(f, g, \pi_x) +\ \delta(g) \qquad 1$$

The explanation of local model x is based on the minimization of a loss function L that depends on the proximity of complex model predictions while the complexity of the model is low. The G in the above equation 1 represents all possible explanations and represents the proximity of around observations that are considered for an explanation. The complexity is based on the number of features that the linear model needs to use. The user can set this complexity parameter. One major issue with using LIME methods is they are computationally expensive and also time-consuming as they need to train a new linear model for every observation.

To reduce this time complexity and maintain the same level of explanations, the one proposed method in this work is to use correlations instead of training local linear models. In this idea, once the black box model makes predictions on the permuted data, the correlations between the features and prediction labels are obtained, and weights are calculated. This correlation approach is a simple yet powerful idea to improve the speed with similar performance. The above equation 1 is modified to reduce time complexity, as shown in below equation 2. In this equation, 'p' is the prediction, and Xi is the feature that needs to be correlated with prediction.

$$Explanation\ (p) = \forall_{i\ \in\ G}(\mathrm{Corr}(X_i, P)) \qquad 2$$

All attribute weights are calculated for each observation to understand their impact. These weights will help us understand whether an attribute supported or contradicted prediction. The positive weights correspond to a supporting attribute, and the negative weight corresponds to a contradicting attribute. As the attribute importance is given to each feature based on the weighted correlation, it is useful to calculate the weights of an attribute over all the predictions to observe their performance globally. To find their global performances, here is the proposed pseudo code. The below pseudo-code is developed to take the tradeoff between supporting and contradicting attributes. To find the global importance of an attribute, we add the weights of each prediction if that attribute supported the correct prediction or it contradicted an incorrect prediction. The assumption is that an attribute that contradicts an incorrect prediction is a good one. In this way, the proposed method calculates the global importance of all attributes and helps understand attributes that highly supported the algorithm predictions and contradicted algorithm predictions.

```
for all rows to be explained do
        {calculate supporting / contradicting values with LIME for the row
         for all those values do
            {if prediction is correct:
                 {if value is positive (supporting):
                      add the value to the weight of the column
                  else:
                        do nothing
                  end if
             else (Wrong prediction):
                  if value is negative (contradicting):
                         add the absolute value to the weight of the column
                  else:
                         do nothing
                  end if
              end if
           end for
end for
```

4.4 Results

The training data set is used in 5-fold cross-validation for training and testing predictive models. The predicted labels for all the samples in the dataset were extracted for confidence calculation and individual student label prediction. The primary performance metrics considered in this comparison were AUC, Kappa, RMSE, and linear aggregation of AUC and RMSE ((1-RMSE) + AUC). We observed that Gradient Boosted Tree performs better with 0.99 AUC and 0.994 kappa compared to all other algorithms. We also find that the KNN algorithm performs better with an AUC of 0.619 and kappa values of 0.32.

Table 4.2: The training set 5-fold cross-validation performance metrics based on feature selection and voting machine mechanism.

Algorithm	AUC	Accuracy (%)	Kappa	RMSE	AUC+(1-RMSE)
Gradient Boosted Tree	**0.995**	**99.78**	**0.994**	**0.0462**	**1.9488**
K-NN (5)	0.619	80.94	0.32	0.436	1.183
Decision Tree	0.508	75.37	0.025	0.496	1.012
Random Forest	0.5	74.94	0	0.5005	0.9995
Deep Learning	0.592	40.68	0.1081	0.772	0.82
AutoMLP	0.5	74.94	0	0.500	1
Logistic Regression	0.5	74.94	0	0.5005	0.9995
Yeung et al. [70]	**1.00**	**-**	**-**	**0.039**	**1.961**

Table 4.2 shows seven machine learning algorithm performances on individual student label prediction once their confidences were calculated and labels were predicted. The metrics of the best-performing model on training data set in our study and previous work [70] were similar with a slight margin.

Table 4.3: Comparing test set performance metrics on trained models based on the voting machine mechanism

Algorithm	AUC	Accuracy (%)	Kappa	RMSE	AUC+(1-RMSE)
Gradient Boosted Tree	**0.834**	**81.97**	**0.619**	**0.424**	**1.41**
K-NN (5)	**0.812**	**87.79**	**0.692**	**0.349**	**1.463**
Decision Tree	0.5	67.44	0	0.570	0.93
Random Forest	0.5	67.44	0	0.570	0.93
Deep Learning	0.572	44.76	-0.329	0.743	0.829
AutoMLP	0.5	67.44	0	0.570	0.93
Logistic Regression	0.5	67.44	0	0.570	0.93
Yeung et al. [70]	**0.694**	**-**	**-**	**0.414**	**1.280**

We applied the models trained in a 5-fold cross-validation on the test dataset to observe the model test performance. We find that GBT and KNN algorithms performed better than other algorithms with 0.834 and 0.812 AUC, 0.619, and 0.692 kappa values. Neural network algorithms failed to distinguish both careers, and one more exciting find is the negative kappa value of deep learning as it shows a sharp disagreement in predictions. The performance of the seven machine learning algorithms on the test set related to individual student label prediction was captured in Table 4.3 below. The GBT and KNN models performed better on test data compared to a previous study [70]. The linear aggregation of AUC and RMSE ((AUC+(1-RMSE)) was used as the primary measure for comparison.

4.4.1 Individual Student Confidence for STEM and Non-STEM

From the above comparison, we observed that the GBT and KNN algorithm performs better in predicting student career choice from optimal feature selection. The predicted and actual labels are extracted, and the percentage of samples that were classified accurately for each student was calculated. These confidence values of GBT and KNN for both training and testing datasets related to optimal feature selection were shown in Figure 4.2.

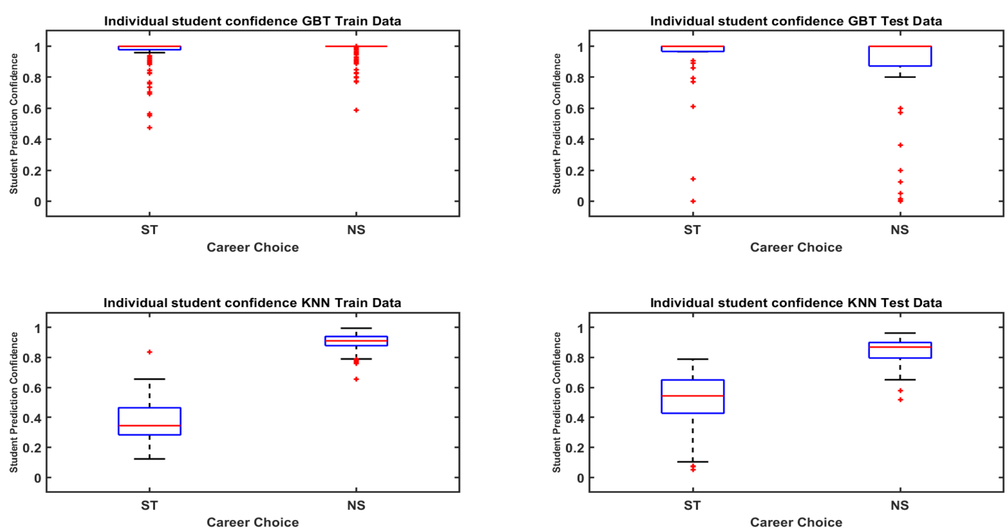


Figure 4.2: Individual student prediction confidence for Gradient Boosting Tree predictions

From the top left subplot related to GBT model training set performance, we observe that the model can predict each student with confidence closer to 1 for both STEM and Non-STEM careers except for some outliers. The top right subplot related to GBT confidence on test data shows that the model can predict better for the test dataset. The bottom subplots belong to the KNN algorithm predictions on training and testing data. KNN is a lazy learner algorithm that does not learn any model from training but memorizes the training set and utilizes the 'k' value to identify nearest neighbors and assign classes based on the majority. By comparing the bottom left and right subplots in Figure 4.2, the confidence of KNN in predicting the test dataset is higher than the training dataset. This is related to good sample distribution for KNN in the test data and the performance fluctuations in the cross-validation folds of training data.

4.4.2 Explain Predictions

The main focus of this study is to understand the predictions made by the adopted machine learning algorithms. For this purpose, we extract all the supporting and contradicting attributes and present the top six in Table 4.4 and Table 4.5 for both accurate and inaccurate predictions. These supporting and contradicting algorithms were classified based on the local Pearson correlation values obtained by calculating the correlation between the attribute and prediction. One should be careful in interpreting support and contradict predictors. For instance, a supporting predictor for a sample with the accurate prediction (Actual Label = Predicted Label) means that this predictor acted positively on predicting the actual label. In contrast, a supporting predictor for inaccurate prediction (Actual Label ≠ Predicted Label) implies that this predictor performed negatively for this prediction. This explanation is similar to contradicting predictors, where the contradicting predictor negatively affects accurate predictions and positive effect on inaccurate predictions.

Table 4.4: Supporting and Contradicting predictors related to GBT model

Accurate Prediction		Inaccurate Prediction	
Supporting	Contradicting	Supporting	Contradicting
NumActions	sumRight	RES_GAMING	NumActions
timeGreater10SecAndNextActionRight	totalFrAttempted	totalFrAttempted	timeTaken
original	sumTimePerSkill	frPast8WrongCount	sumTimePerSkill
frPast5HelpRequest	frPast8WrongCount	hintCount	sumRight
correct	totalFrSkillOpportunities	totalFrSkillOpportunities	totalFrPastWrongCount
manywrong	RES_GAMING	totalTimeByPercentCorrectForskill	Ln

Table 4.5: Supporting and Contradicting predictors related to Deep Learning model

Accurate Prediction		Inaccurate Prediction	
Supporting	Contradicting	Supporting	Contradicting
totalTimeByPercentCorrectForskill	NumActions	NumActions	timeTaken
timeTaken	totalFrAttempted	totalFrAttempted	sumRight
endsWithScaffolding	attemptCount	frPast8WrongCount	hintCount
sumRight	totalFrSkillOpportunities	frTotalSkillOpportunitiesScaffolding	endsWithScaffolding
correct	frPast8WrongCount	attemptCount	hint
sumTimePerSkill	frTotalSkillOpportunitiesScaffolding	totalFrSkillOpportunities	frPast5HelpRequest

4.5 Discussion

Since previous studies focused on extracting knowledge states for each student based on their attempted skills to predict career choice, our focus in this study is to propose a new method that incorporates optimal feature selection on raw data. These features include knowledge and affect states from clickstream records prediction. This process enabled us to extract richer information that can help identify future student career selection. The test set performance metrics for the GBT model in this study shows higher AUC and Kappa compared to the previous research that incorporated DKT+ algorithm to extract knowledge states and used them with the combination of the student profile for career prediction [70].

4.5.1 Findings

From this study, we observed that the performance of the Gradient Boosted Tree algorithm is superior compared to other algorithms. We use AUC and kappa values of both cross-validation and test set to compare model performance. Even the deep learning algorithms and complex neural networks did not perform well in prediction. This phenomenon is due to the capability of GBT to build multiple trees and learn sequentially from previous mistakes during the previous step. The ability of GBT to deal with new data which neural networks fail is one major factor in this result. Also, GBT has a strong regularization function [171][172] compared to neural networks. The ability of GBT to produce more robust predictive models from weak predictors is one main reason for their excellent performance [173].

From the results, we also observe that KNN performed well compared to other complex algorithms. One observation from KNN performance on Cross-validation and testing sets is its ability to predict test set with higher accuracy. We analyzed the reason for low training performance and found that in cross-validation folds, the performances fluctuated between folds, which is why lower training performance. Therefore, it is noteworthy that the generalizability of KNN is limited. The top 5 attributes that have a strong influence on predictions of GBT and the impact of affect states were shown in Table 4.6. The GBT consists of 20 attributes, of which three attributes were related to affect states.

From the ranking mentioned in Table 4.6, student's affect states, knowledge traces, and clickstream records were studied extensively to understand their impact on model predictions. We observe that the "NumActions" have a high impact on overall accurate predictions of GBT but adversely affect the Deep Learning algorithm. This might be due to the differences in the statistical background of algorithms and their regularizations functions. Now in the case of accurate STEM prediction made by the GBT model, attempts and clickstream records support the prediction whereas affect state boredom and disengaged behavior off-task acts negatively on accurate STEM predictions.

Affect state confusion has a high positive influence in predicting Non-STEM class, and disengaged behavior gaming also supports an accurate prediction of this class. Affects states impact on deep learning algorithm seems to be negligible as most of the predictions depend on knowledge states and clickstream records. Gaming the system has a negative impact on STEM career prediction and overall predictions. A previous study by San Pedro et al. also found this relationship between gaming the system and Non-Stem students during their major selection [163]. One reason for the pattern mentioned above might be related to students turning from boredom to off-task, which negatively impacts STEM choice [127].

Table 4.6: GBT variable importance and ranking for model evaluated on optimal features

Ranking	**Attribute**	**Relative Importance**	**Scaled Importance**	**Percentage**
1	NumActions	201489.2	1	89.7
2	sumRight	5085.8.0	0.025	2.2
3	totalFrAttempted	4067.5	0.02	1.8
4	Ln	2244.0	0.01	1
5	hintTotal	1257.0	0.006	0.5
13	RES_Confused	49.8	0.0002	0.02
16	RES_Frustrated	23.4	0.0001	0.01
17	RES_Bored	21.9	0.0001	0.009
19	Correct	12.14	0.00005	0.005

To check the impact of “NumActions” we also looked at the variable importance of GBT in its absence which is presented in Table 4.7. Comparing both Table 4.6 and Table 4.7, we observe that “NumActions” dominates the predictions in its presence. This finding implicates that although the average number of actions (e.g., each action means a learning process of a question) does not reflect the students' STEM choice in the future [70], the individual number of actions influences each student's STEM choice. It means the more the student learns in the mathematical software, the higher the chance the student will choose the STEM career. In this scenario, the boredom of students influences prediction.

Table 4.7: GBT variable importance and ranking for model evaluated in the absence of NumActions attributed

Ranking	Attribute	Relative Importance	Scaled Importance	Percentage
1	totalFrAttempted	15917	1	10.7
2	sumRight	15486.8	0.972	10.6
3	Skill_ID	12447.2	0.78	8.6
4	totalFrTimeOnSkill	10821.2	0.70	7.1
5	Ln	8132.3	0.51	6.4
16	RES_Bored	273.0	0.02	2
19	Correct	122.2	0.01	0.8

Previous studies suggested a high correlation between carelessness and STEM students [69][174]. In this study, the impact of the Average carelessness attribute available in this dataset is investigated to check the model performance based on its presence and absence. With the inclusion of average carelessness, the performance metrics of the GBT model increased. From the predictor explanation, we observe that Average carelessness has a high impact on accurate STEM predictions. This predictor importance is in line with previous studies that proved the importance of carelessness in the case of students opting for STEM fields [147][161][163]. One limitation of this study is related to the use of clickstream data, which depends on multiple factors like time spent on the system, the number of questions answered, which may vary when considering different sets of students that work on the platform during different periods. Another limitation is the generalizability of this study, as the dataset analyzed is from a single platform (ASSISTments), and the predictor relevance is model-specific. Our future work focuses on developing feature selection techniques based on the useful predictors and developing models that efficiently and effectively predict their choice based on their middle school year data. Confusion state has more impact compared to all other affect states.

4.6 Contribution

This study sought to explore the significance of different features, including knowledge states, affect states, and other clickstream features provided by the ASSISTments platform, to predict the individual student's career choice. The developed voting machine mechanism and optimal feature selection techniques demonstrated better prediction performance. The importance of features selected for the best performing algorithm is also analyzed to demonstrate their significance in the prediction models. This study also aims to work with real-world data problems like imbalanced data sets and high correlation attributes. Furthermore, the algorithms based on gradient boosting performs better compared to highly complex neural networks.

However, a significant limitation of the study concerns the mode of the voting machine and the relatively small sample size. The voting machine relies on the trained models based on the samples of students' interaction with the learning system. Therefore, the underlying probabilistic distributions of the samples will significantly influence the performance of the trained models and the voting machines, especially under the imbalance situation. For example, the results of KNN, which heavily depends on the probabilistic distributions of the samples, demonstrated better performance in testing data than training data. Moreover, our work has been validated in mathematical software, ASSISTment, with a sample size of around 500 students, which has potential limitation to be generalized for other science knowledge software.

Our future work will include 1) further develop integrative models to investigating the significance of affective process in learning behavior, 2) transferring the models in other population studies and educational software, such as Blackboard System that has collected plenty of interaction data at the University of Maryland, Baltimore County, 3) exploring interdisciplinary collaboration with other research groups with the ultimate goal of achieving confident models to deliver the affect-sensitive intervention.

Chapter 5: Student-Centered Modeling

With the increasing adoption of Learning Management Systems (LMS) in colleges and universities, research in exploring the interaction data captured by these systems is promising in developing a better learning environment and improving teaching practice. Most of these research efforts focused on course-level variables to predict student performance in specific courses. However, these research findings for individual courses are limited to develop beneficial pedagogical interventions at the student level because students often have multiple courses simultaneously. This paper argues that student-centric models will provide systematic insights into students' learning behavior to develop effective teaching practice. This study analyzed 1651 undergraduate student's data collected in Fall 2019 from computer science and information systems departments at a US university that actively uses Blackboard as an LMS. The experimental results demonstrated the prediction performance of student-centric models and explained the influence of various predictors related to login volumes, login regularity, login chronotypes, and demographics on predictive models. Our findings show that student prior performance and normalized student login volume across courses significantly impact student performance models. We also observe that regularity in student logins significantly influences low-performing students and students from minority races. Based on these findings, the implications were discussed to develop potential teaching practices for these students.

5.1 Introduction

Teaching and learning changed a lot in recent years with the increasing adoption of new computer-based teaching and learning technologies in educational institutions worldwide. As education and learning technology evolves with time, leveraging technological advances to improve teaching practice and student learning will be a prominent research area. The most common technologies used by instructors to deliver course content include Learning Management System (LMS), Course Management Systems (CMS), and Learning Content Management Systems (LCMS) [175]. Even though these systems seem to be synonymous, they have their specific use in the education domain. LMS tools focus on communication, collaboration, content delivery, and assessment, whereas LCMS is similar to LMS with fewer administrative functions. CMS, on the other hand, will focus on the enrollment and performance of students. Of these three systems, LMS is the one that is best suitable for delivering learning strategies to students and is the primary focus of this study.

LMS systems provide a unique opportunity to administrators and researchers to evaluate student data related to time spent on an activity, access times and day, grades, interactions, and many other useful student learning variables. The data logs collected by LMS systems are analyzed with scientific techniques published in the Educational Data Mining (EDM) domain. In their study, Romero and Ventura [176] described that current EDM methods rely on clustering and pattern recognition techniques to categorize students into various groups based on their interaction patterns. Categorization of students using clustering and pattern recognition supports instructors in making changes for a set of students. Teaching practices that impact the entire classroom can be evaluated using predictive analytics that tracks student learning and achievement from the vast amount of interaction data collected by LMS.

Existing research in Learning Analytics (LA) and EDM focused on developing highly accurate predictive models that can estimate student learning outcomes related to assignment scores, course grades, and drop-out probability [177][178]. These course-based predictive models provide early warning to student counselors or instructors associated with a specific course [179][180]. Even with considerable success in this area, many of the student performance prediction models have several shortcomings. One significant issue with course-based models is the bias introduced by teaching style and the type of course (descriptive, programming, mathematical, etc.). This bias impacts these models' scalability across different courses and makes it difficult to understand the student-level factors on their achievement. For example, if a student enrolls in five courses, developing models to study students' progress in these five courses independently is not realistic and gives different insights based on varying features and performances. Therefore, these modeling efforts are limited to reduce different biases introduced by the instructor and the diverse amount of content made available in LMS.

Course-level predictions are suitable for supporting instructor-level decision-making. However, suppose intervention is on student-level behaviors such as study habits or self-regulation skills. In that case, it is beneficial to look at student-centered indicators so that interventions may be more targeted and cost-effective [181][182]. Developing student-centric models that analyze student LMS interactions across courses in a college/university setting will help address course-specific models. This study is the first step in developing models that supports the identification of student-level indicators.

For colleges that have a high penetration of LMS, LMS activity may give a holistic indicator of students' engagement level (behavior engagement specifically). We ask the question to what extent those holistic indicators predict student term Grade Point Average (GPA) performance in the future. We specifically focus on student login-related features to explore this as they can be generalized across courses and act as proxy variables for time management [184][185]. It is also challenging to aggregate other features like discussions, readings, and assessments across courses compared to access-related LMS variables. This study's data is drawn from Blackboard Learn, a commercial LMS software available for colleges and universities to deliver course content and assessments through internet-enabled computer systems.

Most importantly, the data is drawn from all computer science and information systems students at a large public university in the US during the Fall 2019 semester. In addition to student interactions from LMS, we also access demographic and prior student performance data from the university's student administration system to build and interpret downstream predictive models. The university's Institutional Review Board (IRB) approved this study, and all the student-specific demographic and personal information are anonymized by following General Data Protection Regulation (GDPR) standards.

In this work, we focus on model predictions and explanations to understand student learning behaviors. First, we apply new methods to process student interaction data collected across different courses enrolled in a semester to build student-centric performance models based on machine learning principles. Secondly, we utilize a novel approach in local model explanations, correlation, and regression to understand the impact of various features captured by LMS on student performance. One primary reason for using

Locally Interpretable Model Explanations (LIME) is its ability to explain the relationship between predictor variables and predictions, especially the input variable's impact on the outcome. On the other hand, statistical correlation analysis will provide the relation between input predictors and the observed target variable. Correlation analysis does not consider the interaction effect between input variables. We also use a linear regression model to study the output variable's feature importance based on the model coefficients.

5.2 Data & Feature Set

5.2.1 Dataset

For this study, we chose undergraduate student data captured by LMS in Fall 2019 from a large public university in the United States. These students were part of either Information Systems (IS) or Computer Science (CS) departments. The students from these departments were chosen as the instruction format and courses are closely aligned in both of them. The Blackboard system is predominantly used as an LMS to deliver course material, assessment, and grading. The student demographic data captured by a standalone Student Information System (SIS) is used to categorize students based on different demographic variables. A total of 1651 students were enrolled in these two departments in the Fall 2019 semester. Based on student distribution, we categorized students into three ethnicities: White, Asian, and Minority. This study also researches student performance based on their admit types, such as four-year regular student or transfer student. The demographics of student data are provided in Table 5.1. IRB approved this study, and sensitive student data was de-identified based on GDPR standards.

Table 5.1: Student demographics

Demographic	Student Count
Total Students (N)	1651
No of unique courses	440
No of unique course instructor combinations	638
Male : Female	1302 (79%) : 369 (21%)
White : Asian : Minority	630 (38%) : 495 (30%) : 526 (32%)
4 – Year : Transfer	976 (59%) : 675 (41%)
Full Time : Part Time	1446 (88%) : 205 (12%)
IS : CS	934 (57%) : 717 (43%)
1st Yr : 2nd Yr : 3rd Yr : 4th Yr	115 (7%) : 329 (20%) : 515 (31%) : 692 (42%)
<= 3 : 4-5 : >5 (Courses enrolled)	298 (18%) : 1035 (63%) : 318 (19%)

5.2.2 Feature Extraction

We explored various LMS features related to student logins, content accesses, time spent, discussion posts, assignment submissions, and time intervals based on earlier literature. While exploring these features, we identified that only three features could be commonly extracted from different courses: Student Login Counts, Time intervals & prior performance.

One of the significant challenges while building a student-centric model on LMS data is to extract aggregated features that are least biased. As Blackboard's content is dependent on instructor and course, it is crucial to mitigate the variations caused by these factors on aggregate student variables. This work employs multiple statistical measures to reduce these issues. The details are explained in the below sub-sections.

5.2.2.1 Normalized Login Volume

Earlier studies identified that student performance prediction is strongly dependent on the volume of student logins. One challenge with counting the student logins in Blackboard is its inability to find which course they accessed during each login. Also, calculating the total login count introduces a hidden bias as courses with more content on Blackboard prompt students to login more often than other courses with less content and flexible deadlines. To mitigate this issue, our work followed the below steps to extract student login features.

1. Extract all courses enrolled by all students in IS and CS.

2. Count the total number of logins for all students irrespective of their department in these extracted courses.

3. Calculate the Z-scores of student logins in each course. The reason for doing this is to mitigate the bias introduced by variations in the absolute count of logins, as course logins vary a lot between students. Z-scores provide a value that helps understand if student logins are higher or less than average logins in a specific course.

4. Once the z-scores are calculated for all courses. We extract a vector of login z-scores for each student based on their enrolled courses.

5. As predictive models do not take vectors of variable length as input, this work extracts seven significant statistics from the login vector: mean, median, minimum, maximum, standard deviation, skewness, and kurtosis.

5.2.2.2 Login Regularity

Apart from student login volumes, the regularity between logins also provides valuable insights into student achievement as regularity is related to self-regulation capabilities. In this work, we utilize an entropy-based method to extract features that define student login regularity in each course. In information theory, entropy is used to define uncertainty or randomness [186]. Entropy measure will explain if student's logins are regular (less random) or irregular (more random). Based on this concept, if the entropy value is high, then a student has an irregular login pattern, and if the entropy value is low, the student has a regular login pattern. The steps to calculate student regularity features are given below.

1. Extract all course accesses with timestamps for every student in IS and CS.

2. Calculate the difference between timestamps. This difference will give a vector of time intervals for each course enrolled by a student.

3. Calculate entropy using the KL estimator with the k-nearest neighbor method proposed by Kozachenko and Leonenko [186]. KL estimator uses k-nearest neighbor distances to compute the entropy of distributions. The reason for adopting this method instead of Shannon entropy is based on the time interval vector's continuous characteristic [187].

4. Once the entropies are calculated, we get a vector of entropies for each student based on the number of enrolled courses. We then calculate the seven statistics similar to student logins: mean, median, minimum, maximum, standard deviation, skewness, and kurtosis.

5.2.2.3 Login Chronotypes

Studies in chronobiology and chronopsychology showed variation in different individual active periods at different times of the day [188][189]. These studies classify an individual into either morning type or evening type based on their high activity time. For example, if an individual is highly active in the morning compared to the evening, they are considered morning type and vice versa. Inspired by this work in human psychology, this work divides a day into four-time bands T1 (12 AM to 6 AM), T2 (6 AM to 12 PM), T3 (12 PM to 6 PM), and T4 (6 PM to 12 AM) and extract student logins based on these four time bands. In addition to this, this work also extracts the logins on weekdays and weekends to study their influence on student performance.

1. Count the number of logins during each time band and on weekdays and weekends for each course.

2. Calculate the mean of the login count vector for each time band and weekday/weekend.

3. Normalize the login count with the number of courses enrolled by an individual student. This normalization will mitigate the bias introduced by the number of courses enrolled across the student cohort.

This work also utilizes the demographic and prior performance measured by GPA features captured by the SIS system. The prior GPA is a proxy variable for a student's pre-existing characteristics like IQ, motivation, prior knowledge, and metacognitive skills. These features were listed in below Table 5.2.

Table 5.2: Student demographic features

Demographic	Values
Start GPA (Prior Performance)	Cumulative GPA available till the start of the semester
Gender	Male & Female
Ethnicity	White, Asian & Minority
Student Year	Freshman, Sophomore, Junior & Senior
Admit Type	Regular & Transfer
Enrollment Type	Full time & Part time
Student Age	Continuous variable

5.3 Methodology

The methodology section detailed the predictive modeling approach to predict student end-of-term GPA in fall 2019. In addition to this, we also describe the correlation-based LIME method to explain the features that contribute to model predictions. The workflow of developing student-centric models is depicted in Figure 5.1.

5.3.1 Predictive Modeling

This work studied five of the most common regression models for comparison purposes. The selected models include Generalized Linear Model (GLM), Decision Tree (DT), Support Vector Regressor (SVR), Random Forest (RF), and Gradient Boosted

Regressor (GBR). As model hyperparameter influences their predictive performance, we utilized a grid search mechanism to select multiple parameters to predict with high accuracy. We also adopted a feature selection method based on a multi-objective evolutionary algorithm in addition to a hyperparameter search. This feature selection algorithm evaluates each feature set based on Pareto-optimal that balances model complexity and accuracy. The details of models and hyperparameter search criteria are discussed below.

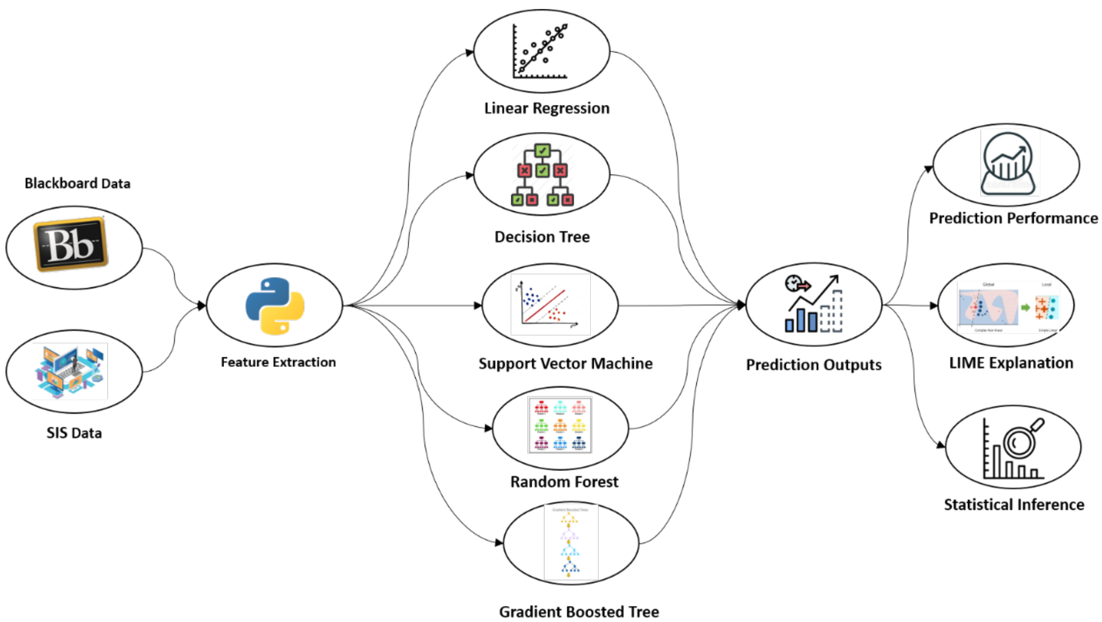


Figure 5.1: Student-centric Model Workflow

Generalized Linear Model: GLM is an extension of traditional linear models that fits input data by maximizing the log-likelihood. The regularization parameter is set so that the hyperparameter search space looks for an alpha value that fits between ridge and lasso regression. An alpha value of 1 represents lasso regression, and an alpha value of 0 represents ridge regression. This study searched for the best alpha value using a grid search between 0 and 1 in increments of 0.1.

Decision Tree: The decision tree algorithm is a collection of linked nodes intended to estimate the numerical target variable. Each node in the tree represents a rule used to split an attribute value. The node uses a least-squares criterion to minimize the squared distance between the average value in a node compared to the actual value. The hyperparameter search space for this algorithm evaluates both maximal depth and pruning. The maximal depth value varies between 1 and 100 in increments of 10. Pruning will make the DT algorithm use multiple criteria like minimal gain, minimal leaf size, and pruning alternatives to decide the stopping criterion.

Support Vector Machines: The SVM used in this study is built based on Stefan Reupping's mySVM [190]. This algorithm will construct a set of hyperplanes in a high-dimensional space for regression tasks. A good hyperplane is decided based on the functional margin. The hyperparameter search space focused on both dot and radial kernel functions with a C (SVM complexity) value range between 10 and 200. The kernel gamma function is set for a radial kernel with a range of 0.005 and 5 with three logarithmic increments.

Random Forest: A RF model builds an ensemble of decision trees on bootstrapped datasets. The splitting criteria are similar to a decision tree. The regression outcome is the average of the observed train data GPA present at that end node. We only tuned the number of trees hyperparameter to reduce the time complexity of the execution. The number of tree searches varied between 10 and 1000 trees in 10 linear steps.

Gradient Boosted Tree: The GBT model builds multiple regression trees in a sequence by employing boosting method. By sequentially applying weak learners on incrementally changed data, the algorithm builds a series of decision trees that produce and ensemble weak regression models. As GBT is a non-linear model, we search hyperparameters related to the number of trees, learning rate, and maximal depth. The number of tree values varies between 1 and 1000 in five quadratic increments. The learning rate varies between 0.001 and 0.01 in five logarithmic increments, and the maximal depth parameter varies between 3 and 15 in three logarithmic increments.

5.3.2 LIME Explanation

The concept of Locally Interpretable Model Explanations (LIME) was introduced to explain the predictions made by black-box models that deal with classification problems. LIME explains each prediction made by a complex model by training a surrogate model locally [170]. However, this earlier methodology is not scalable to deal with categorical variables, tabular data, and regression problems. In this work, we adopt the correlation-based LIME method available in RapidMiner to explain machine learning models' predictions [191][192][193].

1. Perturb data in the neighborhood of each sample in the dataset. The number of simulated samples can be user-defined. A higher number of simulated samples will provide higher accuracy of explanations but at the cost of more run times.

2. Make predictions using the ML model for all the simulated samples around each original sample in the dataset.

3. Calculate the correlation between each feature in the dataset and the target variable.

4. The features that have a positive correlation are considered supporting features, and features with negative correlation with predicted outputs are referred to as contradicting features.

As LIME provides a feature importance value for each feature at each sample, we aggregate the importance value for all samples to build global importance for each variable. The significant advantage of this method compared to traditional global importance methods is its flexibility. As model global importance's are calculated across all samples in the data, the LIME based feature importance can be calculated for subsets of data. This flexibility allows users to understand each feature's role for different sets of populations present in a dataset.

In addition to applying the LIME methodology, this work also studies univariate and multivariate feature importance on student performances by applying correlation and linear regression methods. The student dataset used in this study is divided into multiple subsets containing different student groups based on various demographics. A correlation value is calculated between input features and student end-of-term GPA. This value provides us with an intuition about the impact of multiple features on student performances related to different demographics. As correlation only provides independent variable importance on student performance, we also adopt a linear regression model to explore the variation of feature importance based on coefficient values. Applying a linear regression model will also consider the interaction effect between input features to fit the outcome variable.

5.4 Results

This results section is divided into three subsections based on the three research questions we are focusing on in this study. The first subsection will detail various predictive models' performance on longitudinal student interaction data collected during the fall 2019 semester. The second subsection will detail the importance of student logins and regularity on performance predictions based on LIME methodology. The final subsection will discuss the importance of input features based on correlation and regression methods.

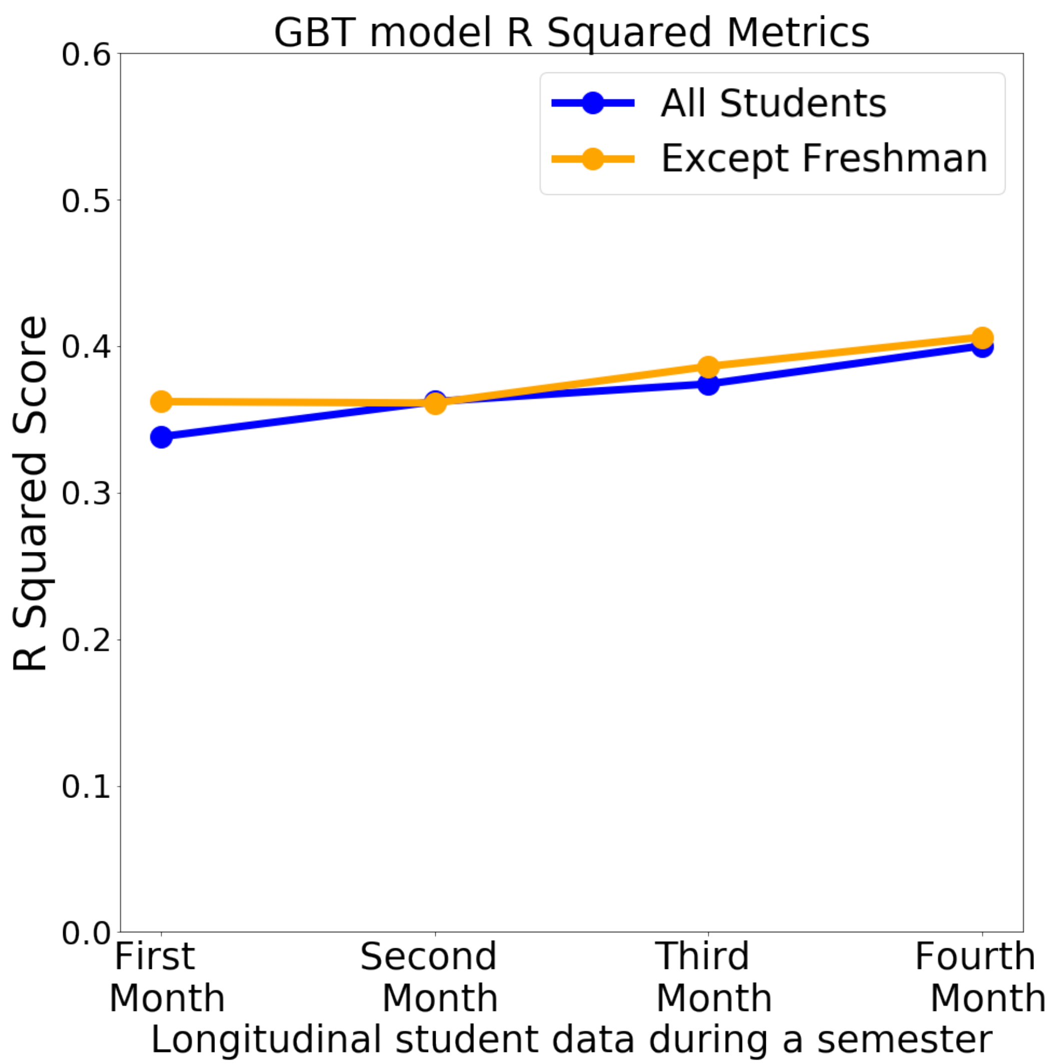


Figure 5.2: Compare performances of GBT model on different longitudinal datasets

5.4.1 How different student-centric machine learning models perform in predicting student end-of-term GPA?

The five machine learning models adopted in this study were evaluated using a five-fold cross-validation method. In this method, the student data is divided into five equal folds at a student level. Four of the five folds are used for model training in every iteration, and one fold is used for model testing. The machine learning models are evaluated based on two performance metrics: R squared (R^2) and Root Mean Squared Error (RMSE). The output performance metrics are the average of five test fold performances.

Table 5.3: Student features from start to end of the first month

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.213	0.249	0.657	0.633
DT	0.266	0.270	0.638	0.633
SVM	0.216	0.324	0.666	0.607
RF	0.332	0.353	0.607	0.588
GBT	**0.338**	**0.362**	**0.602**	**0.581**

This study divided a semester into four parts to understand the impact of longitudinal interaction data across the semester on predictive model performances. This analysis will support the amount of data needed to balance predictive performance and early detection for interventions. The performance metrics evaluated on these four cumulative datasets will help understand the amount of student data needed to make accurate predictions. Table 5.3, Table 5.4, Table 5.5, and Table 5.6 present the machine learning models' results evaluated on four cumulative datasets. While differentiating student performance based on multiple longitudinal datasets, we also study algorithms' performance without Freshman student data. This differentiation is to study the impact of missing start GPA feature values for first-year students as most of the full-time regular students in US universities start in the Fall semester.

Table 5.4: Student features from start to middle of the semester

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.257	0.266	0.67	0.628
DT	0.263	0.295	0.67	0.618
SVM	0.195	0.315	0.705	0.609
RF	0.360	0.352	0.621	0.591
GBT	**0.362**	**0.361**	**0.622**	**0.586**

Table 5.5: Student features from start to end of third month

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.25	0.266	0.644	0.626
DT	0.255	0.255	0.658	0.650
SVM	0.335	0.344	0.612	0.597
RF	0.371	0.386	0.589	0.575
GBT	**0.374**	**0.386**	**0.588**	**0.572**

Table 5.6: Student features from start to end of semester

Model	R^2		RMSE	
	All Students	Except Freshman	All Students	Except Freshman
GLM	0.251	0.269	0.644	0.625
DT	0.246	0.274	0.657	0.641
SVM	0.320	0.289	0.616	0.627
RF	0.387	0.410	0.585	0.564
GBT	**0.400**	**0.406**	**0.575**	**0.562**

From the above tables, we observe that the GBT model performed better than the other four models based on the tradeoff between R squared and RMSE values. We also observe that there is no significant difference in student end-of-term GPA prediction with and without freshman details. This might be due to less sample size (7%) related to the freshman cohort. From Figure 5.2, it is also evident that there is a gradual increase in the performance of the GBT model as we add data to predictive models as the semester progresses. Even though there is an increase in performance if we add all data captured during the semester, it doesn't help much for real-world interventions as activities that effects student performances will be completed by the end of the semester. Based on this understanding, we focus on data captured until the middle of the semester for feature importance study.

We also validated the best performing model (GBT) developed on Fall 2019 middle of semester data by testing them on Spring 2019 and Fall 2018 student data related to IS and CS domains. The results in Table 5.7 shows that student-centric models are extendable across different semesters. Additionally, this work tested the model transferability by validating them on student data from other departments in the Fall 2019 semester. From Table 5.8, we can observe that the test performance of the model developed on IS and CS student data showed similar performance metrics on students enrolled in four other departments.

Table 5.7: Validating the extendability of models developed on Fall 2019 student data to different terms

Term	R^2	RMSE	Student Count (IS & CS)
Fall 2019 (Main Model)	0.362	0.622	1559
Spring 2019	0.370	0.683	1545
Fall 2018	0.360	0.722	1486

Table 5.8: Validating the transferability of models developed on IS & CS student data to other departments

Department	Degree	R^2	RMSE	Student Count
IS & CS (Main Model)	BS	0.362	0.622	1559
Biological Science	BS	0.381	0.675	491
Bio Chemica & Molecular Biology	BS	0.381	0.737	226
Mechanical Engineering	BS	0.385	0.657	391
Psychology	BS	0.317	0.849	215

5.4.2 How do student login and time interval patterns across courses influence student learning outcomes?

To answer research question 2, we adopted a stepwise feature addition study that inputs features by adding one by one into the model and evaluates the performance based on R square and RMSE values. This study is performed on student data collected until the middle of the semester as models developed during this stage will help identify student-level indicators and give enough time to deploy interventions that improve student performance. We first start with inputting student Start GPA (Cumulative GPA till the start of Fall 2019 semester) as start GPA showed a high correlation with end-of-term GPA based on our preliminary analysis. We then add normalized login volumes, login regularity, and login chronotypes in a step-by-step method. Figure 5.3 shows the R squared performance metric of student-centric models with different input variables.

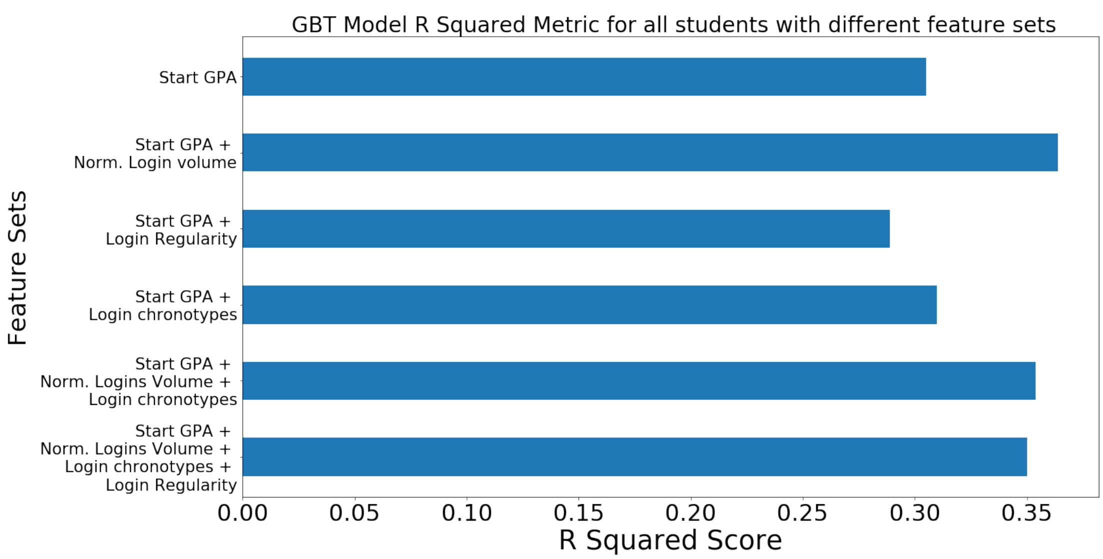


Figure 5.3: Compare performances of GBT model on different input feature sets

From Figure 5.3, we observe that student's start GPA with normalized student login volumes across courses adds more predictive power to machine learning models. This observation is also supported by earlier studies [194][195] that showed the importance of student login counts on student course grades and score predictions. Another observation is related to the importance of adding student self-regulation capability based on login regularity measured using entropy statistics. Based on Figure 5.3, we observe that adding login regularity features with student login features and start GPA adds slightly more predictive power compared to the model with only login regularity and start GPA features. In addition to these observations, we also observed that login counts based on login chronotypes with start GPA did not add much predictive power to machine learning models. We also imply that student aggregated login volumes might be adding the same information as login chronotypes from these results.

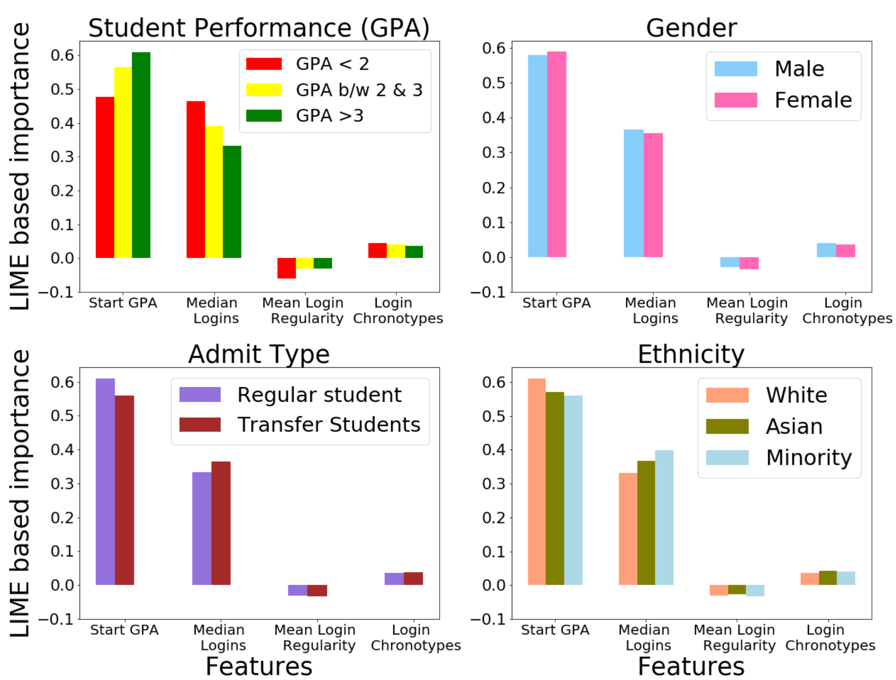


Figure 5.4: LIME importance's for different student groups divided based on GPA, ethnicity, admit type and gender.

5.4.3 Is there a significant variability in feature importance for students coming from diverse demographics?

One limitation of using the earlier mentioned model-based feature importance study is its inability to explain each feature's importance on different student cohorts. To address this issue and understand the importance of login volumes and regularity features on different student groups, we adopt three approaches: one based on LIME, the second based on correlation analysis, and the third based on linear regression.

5.4.3.1 LIME based Importance's

LIME-based approach extract feature importance at the local level also called local fidelity. By applying the LIME method explained in the methodology section, we extract feature importance's for different student groups categorized based on their demographics.

From Figure 5.4, we can observe that cumulative student GPA at the start of the semester is an important feature to predict student end-of-term GPA. Student login volumes are the second important feature set for model predictions on different student demographics. This study's focus is also on student self-regulation capability measured by the regularity of logins (entropy). We observe that for students with GPA values less than 2, the regularity of the logins feature played a key role compared to a student with a higher GPA. This observation also holds for students from minority ethnicity. One implication from these observations Is that introducing teaching practices that guide LMS use and time management will significantly impact students with low GPA and a minority race. Start GPA played a slightly less significant role in transfer students than regular students as transfer students join in different years. Their cumulative GPA might not be available at the start of the semester, similar to freshman.

Even though there is a huge imbalance in the number of male and female students present in the dataset, we do not observe any significant difference in feature importance between these two genders. One limitation of the LIME method is related to global importance. The importance’s showed by LIME at the local level do not necessarily correspond to global importance’s. Based on this limitation, we can infer which feature is essential for different students' groups but not quantify them as the importance’s calculated in this study are the aggregate of importance’s provided by LIME for each student.

5.4.3.2 Correlation-based Feature Importance’s

As earlier feature importance methods showed a significant impact of login volumes and login regularity measured by entropy statistic to predict student performance, we adopt Pearson correlation statistic to infer this relationship for different student groups. To do this, we create subsets of student data based on different groups: student GPA, gender, ethnicity, and admit type.

From Figure 5.5, we observe that the student logins count and regularity in logins are highly significant for a student with a GPA lower than 2. We can also observe that as the entropy increases, the GPA reduces. This observation holds true as regularity in student logins represents their self-regulation capabilities. Earlier research showed that students with good self-regulation capabilities perform better in class [196][197]. For other student groups divided based on gender and admit type, there is no significant variation in the importance of logins and entropy on student performances.

Even though the absolute values of correlation observed in Figure 5.5 are not very strong, the comparison between different groups helps understand which features are significant for students from different demographics. In addition to this, we also observe a similar pattern in LIME-based importance's discussed in earlier sections. We can infer that LIME based method also scales well for global feature importance in this study.

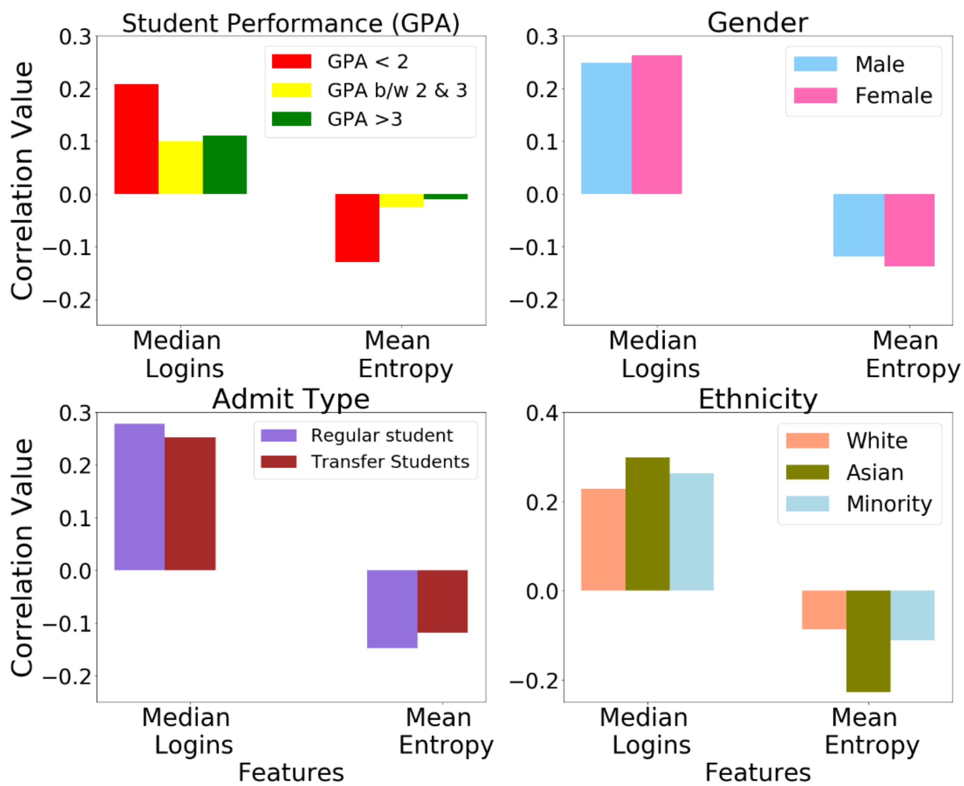


Figure 5.5: Correlation values for different student groups divided based on GPA, ethnicity, admit type and gender.

5.4.3.3 Regression Modeling for Feature Importance

One significant limitation of earlier methods is their inability to capture interaction effects as feature importance might change in the presence of other features. To study the interaction effects, we apply a linear regression model on different categories of student login data collected till the middle of the semester. These student categories were divided based on GPA, gender, admit type, and ethnicity of students. Even though linear regression models are applied on all features discussed in earlier sections, we only report the coefficients of median login volume and mean login regularity in Table 5.9, as these variables are the focus of this study. From Table 5.9, we observe that login volume and login regularity features follow similar directions for students with lower GPAs and students from minority ethnic backgrounds as observed in the LIME and correlation-based analysis. There are some discrepancies in other observations as there is no statistical significance (high p values) for coefficients in these cases. Another reason for focusing on a student from these two groups is their higher attrition rates found in earlier studies [198][199]. Studying these groups closely will help develop targeted interventions in the future.

Table 5.9: Regression coefficients (Significance marked with *)

Student Demographic	Student Groups	Median Logins Coefficient	Mean Login Regularity Coefficient
GPA	**GPA <= 2**	**0.171***	**-0.398***
	GPA >2 & <= 3	-0.013	0.200
	GPA >3	0.065	-0.002
Gender	Male	0.135	0.130
	Female	-0.021	-0.157
Admit Type	Regular	0.399	-0.004
	Transfer	0.611	0.191
Ethnicity	White	0.204	-0.029
	Asian	-0.085	0.201
	Minority Race	**0.201***	**-0.153***

5.5 Discussion

There is a growing interest in building models that capture student behavioral patterns while using LMS systems to predict their performance. Earlier research showed that building efficient models based on LMS data to predict student performances is not a simple task as multiple learning and demographic factors impact student learning processes. Although earlier research in EDM and LA tried to address different issues related to student performance tracking, there is still a gap in developing models that accurately predict overall student performance and explain underlying factors that improve their academic performance. As a step in this direction, this study presents a student-centric modeling approach based on aggregated LMS features to predict and explain the reasons behind varying student performances. This context is both relevant and timely given the increase of LMS adoption and a need for efficient and interpretable model development.

5.5.1 Key Contributions

One primary contribution in this study is the development of student-centric models on aggregated student LMS login data that are least biased towards the diverse course contents and instructor teaching styles. Using the feature extraction methods developed in this study, we were able to build an efficient GBT model that can predict student end-of-term GPA with an average R squared of 0.37 across the semester. Furthermore, models built at different durations of a semester showed only a slight improvement in predictive performance after crossing a specific duration (middle of the semester). This observation helps develop models in the middle of the semester to estimate student performance accurately.

In addition to developing student-centric models, this study also focused on understanding the impact of various LMS features on student performances. Earlier studies in this domain primarily focused on the volume of logins. This work also studied the impact of login regularity measured by entropy statistics on student performance by implementing LIME explanation, correlation, and linear regression methods. From our interpretation studies, we observed that students who login regularly into the LMS system have a positive relationship with performance improvement. The model explanation outcomes on this data also showed a positive relationship between increase in student login volumes and GPA. Even though this observation is accurate across all students, it has a slightly higher importance in students from minority races and student with GPA lower than 2 based on the data analyzed. However, we should be careful in coming to solid conclusions based on these outcomes. LMS data is only a snapshot of student learning activity, and many outside factors might influence this. This specific observation prompts us as researchers to look beyond LMS data and collect qualitative data from students related to external factors like economic condition, neighborhood situation and technology availability to understand what impacts their logins.

We also found no significant difference in the impact of LMS features on Male and Female students. This observation is valid as LMS features used in this study are captured objectively rather than subjectively. This observation also holds for regular and transfer students.

Our study also extracted student interaction features based on chronobiology and chrono psychology concepts to understand if there is a student performance variation based on different chronotypes. From the results, we observed no significant difference in performance. The impact of these features is negligible in the presence of aggregated student login volume.

5.5.2 Applications & Limitations

Student performance tracking is a complex process as it depends on multiple dimensions and facets. Developing student-centric models to predict student performance models helps student counselors and educational administrators design student-level interventions that attract students' attention. Also, developing predictive models that estimate students' overall performance in the middle of the semester will make them aware of their predicted end-of-term performance. These predictions might act as an external intervention to improve their performance in the remaining part of the semester. By understanding the difference in LMS features on students from different demographics, researchers and administrators can build more personalized instructional methods suitable for diverse student cohorts.

There were also some limitations in this study. The predictive performance achieved by using aggregate features across different courses enrolled by students is moderate at best. It would be more helpful to explore ways to improve the performance of these models. One possibility is to add other features that target independent content access durations, mid-semester assessments, and other external factors. One major challenge that needs to be addressed in our future studies is to find an effective method to aggregate content level features across different courses enrolled by a student. The dataset used in this study is extracted in a single semester and students from two closely related departments.

To conclude, we built student-centric models to predict student performances that support the development of student-level interventions. We then use the LIME explanations to study LMS features' importance on student performance prediction. Finally, we study the univariate and multivariate feature importance’s using correlation and regression methods and assess them with the feature importance’s extracted in the LIME method.

Chapter 6: Chronotypes, Login Behaviors and Academic Achievement: A Causal Analysis of Their Relationships

Understanding causal relationships between variables help identify the variable that causes a change in another variable. Identifying this relationship is essential to study intervention effects that determine the impact of external force on the possible outcome. In this work, we focus on learning the causal relationships between student learning behaviors encoded in their login data captured by Learning Management Systems (LMS) and their performance outcome at the end of the semester. Additionally, this work also identifies student behaviors based on different chronotypes characterized from their login activity. The study data is collected from 1688 undergraduate students enrolled in Information Systems and Computer Science departments at a public research US university in the Fall 2019 semester. This study's findings showed a significant causal relationship between student login volume on the LMS system, prior performance, and their end-of-term GPA. In addition to this, we found that the impact of student chronotypes on academic performance discussed in earlier course base studies doesn't scale well to student-level models.

6.1 Introduction

Understanding student learning behaviors is a complex process as it is affected by a multitude of factors like self-regulation, instructional design, and social environment [1]. However, it is not efficient to study all the influencing factors as some factors are not feasible to intervene compared to others. Some have a higher impact on the performance at an individual level than others. Earlier research in education identified self-regulation as a highly impactful and intervenable factor [200][201]. Self-regulation, in general terms, is defined as one's ability to control behavior, thoughts, and emotions to achieve their goal. In learning, self-regulation assists students in managing their behaviors, thoughts, and emotions to drive their learning experiences successfully.

Studying student self-regulation capabilities is an arduous task as it involves multidimensional and complex factors like planning, goal setting, time management, and self-monitoring. Many of these factors are hard to quantify without student self-reports or other psychometric evaluations, and these reports are prone to biases like social desirability and reference bias [202][203]. To mitigate these biases, researchers in education focused on extracting data from computer-based systems like Learning Management Systems (LMS) as they are used to deliver course content and have the capability to capture student interaction and assessment behaviors non-invasively [87][204]. Student interaction data with LMS proved to be a valuable resource in identifying learning strategies and track patterns among students that strongly support their academic performance. LMS systems also enabled researchers to study time management strategies, a component of self-regulation adopted by students [183][205]. In contradiction to traditional time management studies that utilize self-reported questionnaires, student time management strategies are analyzed based on login and submission patterns in LMS systems.

Earlier studies focused on the relationship between self-regulation and student performance through LMS systems used student interaction features as proxy variables for self-regulation [7][206]. These studies showed a significant correlation between student login patterns and their academic performance. However, it is still challenging to design interventions as correlations do not necessarily mean a meaningful cause and effect between student login behaviors and academic performance. In addition to this, very few studies focused on the relationship between time management strategies identified in LMS systems with the biological nature of human functionality. Therefore, these earlier modeling efforts are limited to predicting student performance as a function of their login patterns and fell short of identifying and recommending helpful strategies that support student self-regulation components like time management that improves student academic improvement.

In the context of student learning in engineering and computing research, multiple studies focus on student self-regulation and time management strategies. These studies focus on student behaviors related to their time compliance with assignments, including procrastination behaviors and ways to support the earliness of student work [207][208][209]. These earlier studies showed that early starters have better learning outcomes, but there is little emphasis on what times these students work during a day and how these learning time patterns affect their performance. To address the research gap discussed earlier, we explore the below two research questions.

RQ 1 How do different student learning times (chronotypes) impact their academic performance?

RQ 2 Is there a causal relationship between student login behaviors, chronotypes, and their academic performance?

This work focuses on studying login time patterns, also referred to as student chronotypes, to understand the relation between student learning times and their impact on performance outcomes. In addition to this, we also study the causal relationships between student learning behaviors encoded in LMS login variables and their performance outcomes. To perform this study, we extract hourly student login volume data and perform clustering to observed if any visible patterns exist among students. We then utilize these cluster outcomes to develop predictive models to study the impact of student chronotype on their performance. Secondly, we adopt student login variables from an earlier study and combine them with student chronotypes to understand causal relationships between student behaviors and their performance outcomes. One primary reason for studying causal relationships is to understand if there is a direct cause-and-effect relationship between login variables and student performance. This understanding will support the development of intervention techniques that enhance learning and support student academic achievements.

6.2 Data & Features

In line with our previous studies, we capture student interaction data from the blackboard LMS system during the Fall 2019 Semester. These are on-campus students who enrolled in undergraduate level Computer Science and Information Systems majors. The student demographic data is accessed from the Student Information Systems (SIS) and used to classify students based on demographics. We adopt similar features from our previous study [210], focusing on student demographics and normalized login variables. In addition to this, to study the student chronotypes, we extract hourly student login volume data till the middle of the semester. These hourly login volumes per student are then used to calculate the hourly percentage of logins at the student level. This calculation of percentages is to normalize login volume that varies between different students based on the course they were enrolled in and the varying content on the blackboard for each course. The detailed features adopted in this study are given in Table 6.1 below.

Table 6.1: Features adopted to study chronotypes and causal relationships.

Feature	Description
Demographics	Age, Gender, Student Year, Type of Admit, Full time/Part time, Student Ethnicity
Enrolled Courses	The total number of courses enrolled by student in Fall 2019
Student Start GPA	The cumulative GPA of student till the start of Fall 2019 Semester
Normalized Login Volume	Statistics related to login z scores of students per each enrolled course: Mean, Median, Standard Deviation, Minimum Score, Maximum Score, Skewness, and Kurtosis
Student Regularity in Logins	Statistics from KL Entropy of student login intervals per each enrolled course: Mean, Median, Standard Deviation, Minimum Score, Maximum Score, Skewness, and Kurtosis
Weekday & Weekend volumes	Login counts on weekdays and weekends
Hourly Login Volume Percentage	Count cumulative login volumes per hour and normalize them by calculating the percentage of logins per hour.

Hourly Time spent	Count the aggregate time spent per 60 minutes every hour till the middle of the semester and then calculate the percentages by normalizing across 24-hour bands in a day.

6.3 Methodology

The methodology section of this study is divided into two subsections. In the first subsection, the focus is on studying the chronotypes of students present in the dataset and the impact of these chronotypes on student academic performance. This chronotype analysis pipeline is detailed in Figure 6.1. The second subsection focuses on methodology to study the causal relationships between student login behaviors, chronotypes, and academic performance outcomes.

6.3.1 Chronotype Analysis

This work adopts a clustering-based method to study chronotypes in students. In order to achieve this, we extract the count of student's logins into blackboard systems every hour in a 24 hour day. These hourly counts were then aggregated till the middle of the semester for each student, and then a percentage of logins for every hour based on total logins per student are calculated. To understand the chronotype patterns encoded in student interactions, we utilize the hourly student login volume data to develop cluster models that segregate students based on their varying login patterns. In this work, we adopt the X-means clustering algorithm developed by Pelleg and Moore [211].

X-means algorithm utilizes heuristics to determine the appropriate number of centroids needed to cluster the data. The algorithm requires the minimum and the maximum number of centroid values. This algorithm starts clustering with the minimum number of centroids and then iteratively exploits if using more centroids makes sense according to the data. Whether a cluster needs to be divided into two sub-clusters is determined by the Bayesian Information Criterion, The algorithm also balances the tradeoff between model complexity and precision.

Using X-means over k-means is based on its ability to scale for computationally intensive tasks and accommodate a range of centroid values rather than a fixed, predetermined value needed by k-means. This algorithm is also helpful in avoiding local minima. Another critical input required by the X-means algorithm is related to the choice of distance calculation algorithm. In this work, we adopt Dynamic Time Warping (DTW) for distance calculation.

This study adopted DTW for distance calculation in the X-means algorithm based on its ability to analyze the time component in hourly student login volume data. Traditional Euclidean distance takes pair of data points and compares them with each other, whereas DTW calculates the distance between all data points to enable a one-to-many match [212][213]. Euclidean distance also assumes that the time series are of equal length and calculated distance between corresponding points. This method of corresponding distances doesn't consider shapeshifts in time series data. On the other hand, DTW accommodates shapeshifts by calculating a matrix of distances between different points in time across signals. This is shown in Figure 6.2. From this figure, we can observe that Euclidean connects point to point, but DTW calculates point to point and also points where there is a shift in signal shape. This assumption introduces errors if the data points have shifted between each other. Based on the strengths of DTW over Euclidean distance, we cluster the hourly student login volume percentage using X-mean with DTW as its distance calculation algorithm.

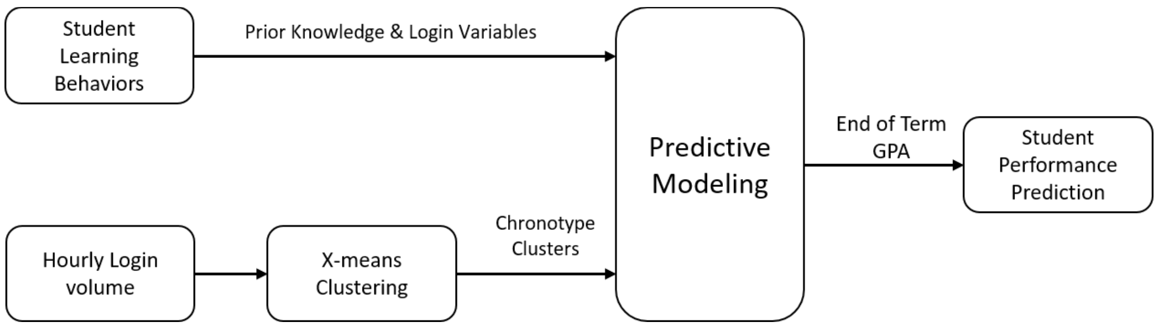


Figure 6.1: Analysis pipeline to study student chronotypes and their impact on student academic performance

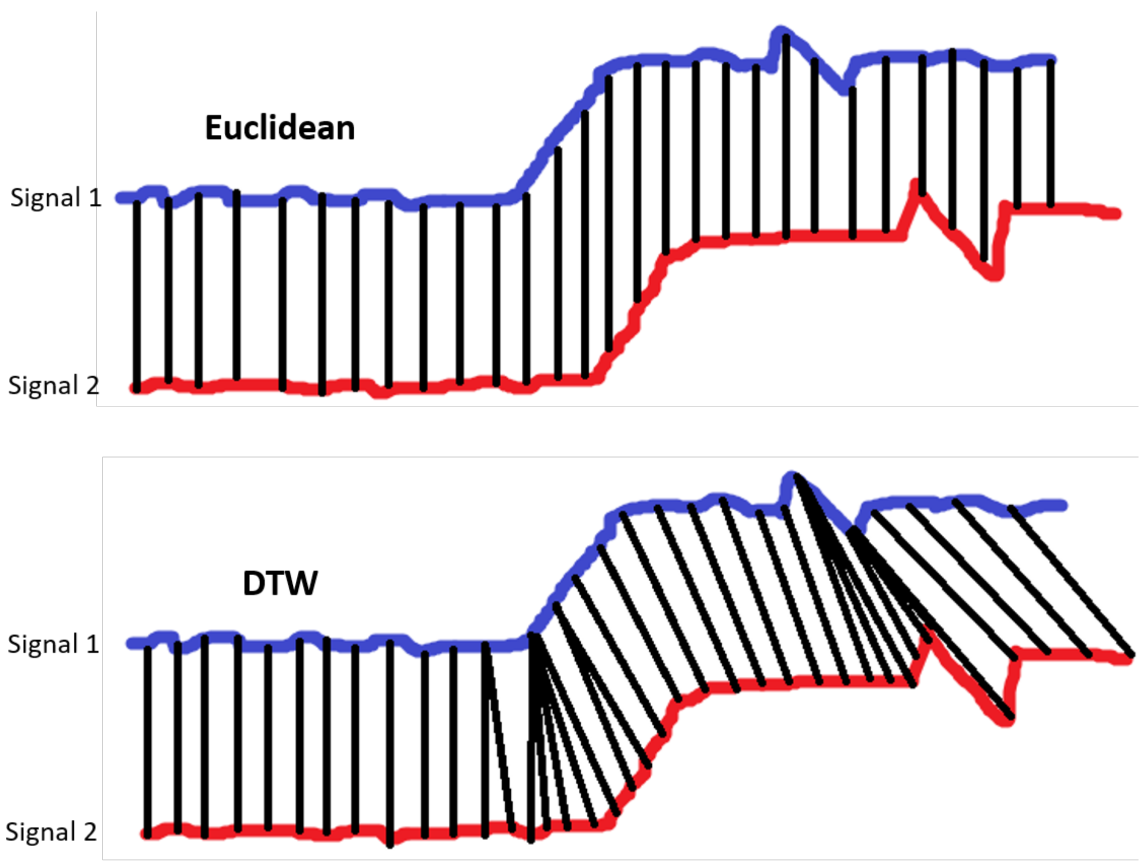


Figure 6.2: Euclidean distance calculation Vs DTW distance calculation

6.3.2 The Impact of Chronotypes and Login Behaviors on Academic Achievement

Our previous study [210] developed predictive models to study the impact of login variables on student academic performances. In line with our previous work, we adopt predictive modeling techniques to study the effects of chronotypes on student performances. To do this, we use the cluster outcomes of student chronotypes and add them as input features detailed in Table 6.1 to predict end-of-term GPA. The selected models include Decision Tree (DT), Generalized Linear Model (GLM), Random Forest (RF), Support Vector Regressor (SVR), and Gradient Boosted Regressor (GBR). We use the same hyperparameter settings from our previous work as listed in Table 6.2. The modeling pipeline, in this case as well utilized the multi-objective evolutionary algorithm to select appropriate features.

Table 6.2: Hyperparameter search space to study the impact of student chronotypes on performance prediction

Model	Hyperparameters
GLM	Regularization set with an alpha value between 0 (Ridge) and 1 (Lasso)
DT	Maximal depth between 1 and 100 in increments of 10. Pruning is set
SVR	Dot and Kernel function with a C (Complexity) value between 10 and 200. Kernel gamma for radial between 0.005 and 5
RF	Number of Trees between 10 and 1000 (10 linear steps)
GBR	The number of trees between 1 and 1000 (5 quadratic increments). Learning rate

	between 0.001 and 0.01 (5 logarithmic increments). Maximal depth between 3 and 15 (3 logarithmic increments).

6.3.3 Causal Relationship between Student Login Behaviors, Chronotypes, and Academic Performance

One of the primary objectives of this study is to understand any causal relationships between student login behaviors and their performance outcomes. This understanding will support the development of efficient and effective intervention techniques that positively impact student academic performance. To do this, we adopt the processing pipeline shown in the below Figure 6.3.

The process shown in Figure 6.3 takes three inputs: student login variables, demographics, and chronotype clusters. In login variables, we specifically focus on normalized student login volume, regularity of student logins, and weekday/weekend logins described in Table 6.1. As these features are a group of independent statistics (mean, median, standard deviation, minimum, maximum, skewness, and kurtosis) grouped into sets based on their relevance, we input them to Sparse Multiple Canonical Correlation Analysis (CCA) algorithm. This algorithm will help learn sparse representation for each feature group by maximizing the correlation among these feature sets with multiple features [214]. This CCA approach results in a composite variable for each feature group. These are represented by a linear combination of small feature sets and are used as inputs for causal structural discovery and inference.

Student demographics are used in two ways. In the first method, we feed these demographics directly to the causal discovery model to study their relationships with student performances. In the second method, we use this demographic information to separate students based on their demographics and explore their causal relationships. Finally, we use chronotype clusters to study the causal relationship between chronotypes and student performance.

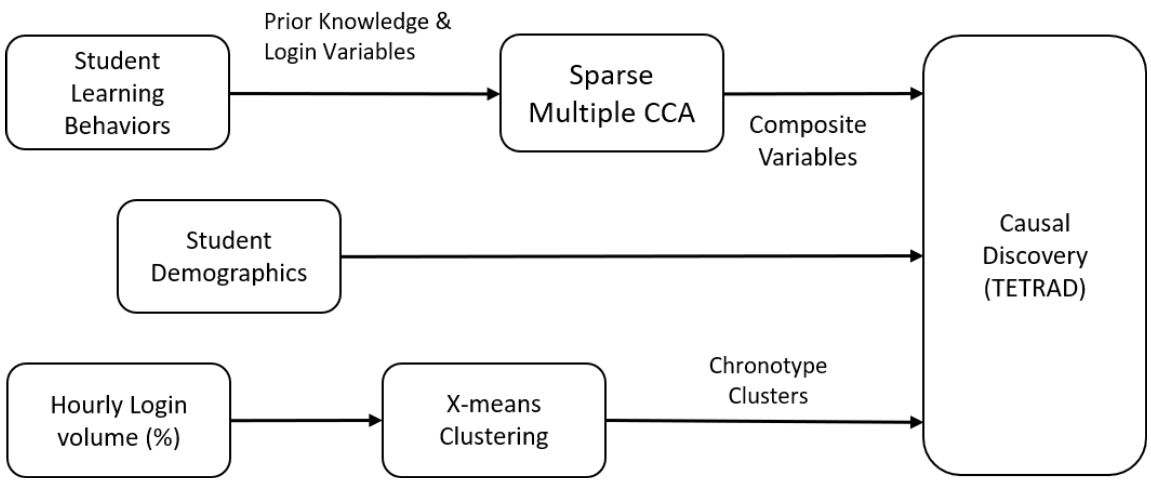


Figure 6.3: Analysis pipeline to study causal relationships between student learning behaviors, chronotypes and performance

6.3.4 Sparse Multiple Canonical Correlation Analysis & TETRAD

This work's causal discovery and inference are performed using a software suite named TETRAD [124]. TETRAD was developed 20 years ago as a drag and drop suite of procedures to explore causal relationships in the input dataset. This suite can take continuous, categorical, mixed, covariance, and correlations as input for causal discovery. This suite consists of multiple proven algorithms that are selected based on the type of data that is inputted for causal discovery. This work adopts two algorithms: Greedy Fast Causal Inference (GFCI) and PC algorithm [215][216]. These algorithms were selected based on their ability to accept prior knowledge that is useful to control the directionality of causal relations.

The GFCI algorithm only works with continuous variables and can output Partial Ancestral Graphs (PAG). These PAG's are causal networks that include hidden confounder variables. The working of this algorithm is based on the combination of two algorithms named FGES and FCI [217]. The FGES algorithm takes input sample data and background knowledge to score using a greedy search mechanism and applies it to a larger sample, and selects the Causal Bayesian Network (CBN) with a higher score. On the other hand, FCI is a constraint-based algorithm with similar functionality to FGES. In addition to this, it entails the set of conditional independence relations that will be satisfied at population-level data. FCI algorithm has two phases names as adjacency phase and orientation phase. The adjacency phase starts with an undirected graph and performs a sequence of conditional independence tests to filter edges between two adjacent variables that are independent. As the output of the adjacency phase is an undirected graph, the directionality between adjacent sets is provided by the orientation phase. In the orientation phase, the stored conditional settings used to remove the adjacencies to reduce edges are used to determine directionality. GFCI algorithm uses FGES to improve both adjacency and orientation phased of FCI by providing a more accurate initial graph for causal network development. The GFCI algorithm in this work is only used to study the causal relationship between login variables and student performance, as this algorithm doesn't accept demographic variables due to their categorical nature.

GFCI is a robust causal inference algorithm for continuous datasets. However, it is unable to handle categorical variables like student demographics. We adopt a PC algorithm to explore causal relationships with both categorical and continuous variables to mitigate this issue. PC algorithms utilized a pattern search method that assumes the underlying structures in data are acyclic. This algorithm also assumes that no two variables have the same latent variable. The only drawback of the PC search algorithm is its inability to show confounding relations in its data. The earlier GFCI algorithm mitigates this issue. This is also one of the primary reasons to adopt two algorithms in this study instead of one for causal structure discovery and inference. Once the selection of causal algorithms is made, we move to structure input data that the TETRAD can access.

The login variables explored in this study can be grouped into multiple sets of features with meaningful roles, as shown in Table 6.1. However, it is not feasible to input a group of variables to represent a single role in TETRAD. This work generates a composite variable from the group of features for every feature subsets using Sparse Multiple Canonical Correlation Analysis (mCCA) [214] to mitigate this issue. The traditional CCA method takes two matrices as inputs to generate a linear combination of variables in each feature set with a high correlation between the two feature sets. This is similar to the way Principal Component Analysis (PCA) works, but on multiple datasets at a time. In addition to this, CCA also maximizes the correlation between datasets while generating composite variables. Sparse CCA is an extension to traditional CCA, where sparsity constraints are imposed. This makes it more compact and yields a more interpretable representation of the data. This study adopts the mCCA function from the PMA package in R. The standardized student login variables were inputted to the mCCA algorithm. A grid search is performed to select the model with the best total correlation and compactness. This model is used to outputs weights for each feature and then generate composite variables for each feature set.

These composite variables are finally inputted to GFCI or PC variant algorithm to perform causal structural discovery and inference.

6.4 Results

The results section is separated into two subsections to address the two research questions discussed in this study. The first subsection reports the clustering results related to student chronotype analysis. The second subsection reports the results obtained by analyzing causal relationships between student login variables, chronotypes, and performance outcomes.

6.4.1 How do different student learning times (chronotypes) impact their academic performance?

We cluster the student hourly login percentages using X-means with DTW distance calculation to study different chronotypes. The results of the clustering algorithm are shown in Figure 6.4. For this Figure 6.4, we can observe that the X-means algorithm found three clusters to be optimal for the input data. The x-axis in Figure 6.4 shows the hours in a day based on the 24-hour clock. It starts with H1 representing 12 AM to 1 AM and ends with H24 describing 11 PM to 12 AM with 1-hour incremental time bands. The y-axis in Figure 6.4 represents cluster centroids for each corresponding time band and cluster.

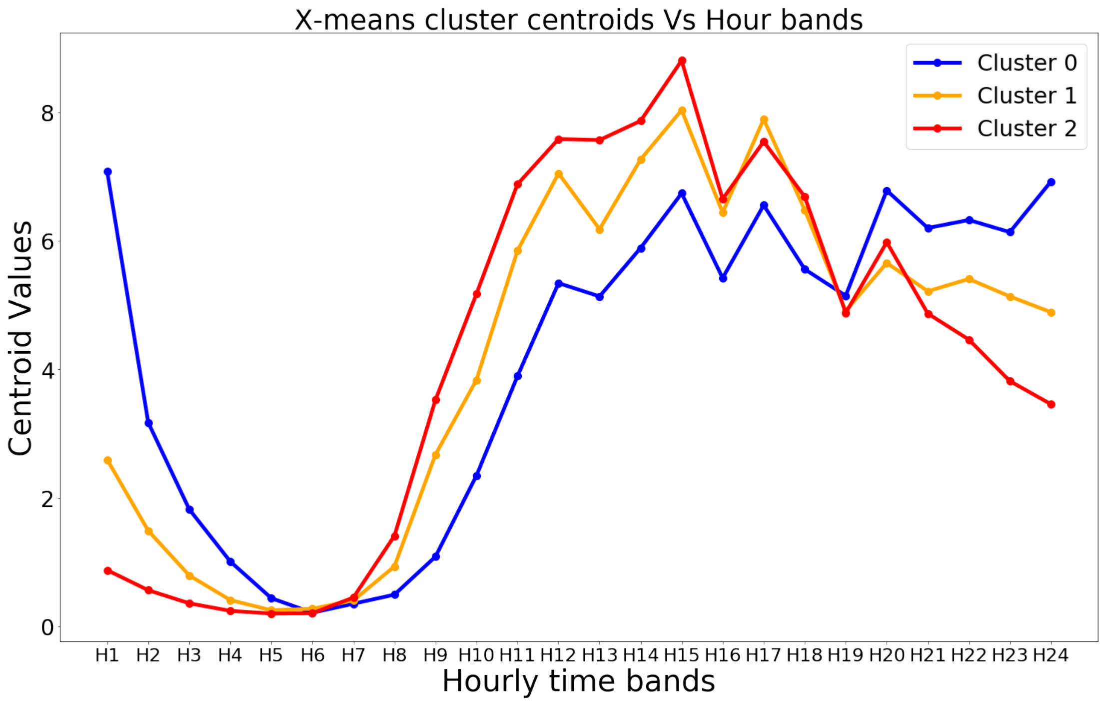


Figure 6.4: Clustering outcome of normalized hourly login volume data to study student chronotypes

From cluster 0, we can observe that students' login percentages slowly start to increase from the morning at 8 AM (H9) and reaches a peak between 12 AM and 1 AM (H1). We can classify these students as most active in the evening with peak activity at night. There are a total of 239 students clustered into this group. Cluster 1 in Figure 6.4 starts to see an activity increase from 8 AM and reaches a peak activity at 2 PM and 4 PM and suddenly drops till 7 PM in the evening and gradually stabilizes there till 11 PM and then dips during the nighttime. There is a total of 560 students in cluster 1 that are assigned to this pattern. In final cluster 2, consisting of 889 students, the pattern is similar to cluster 1, but it doesn't stabilize at 7 PM. The dip in activity of these students continues with very few of them active at night. Based on the patterns found in these three clusters, we can classify cluster 0 as "Afternoon to Night," cluster 1 as "Afternoon to Evening," and cluster

2 as "Active Afternoon." Even though these three clusters have some similarities in the afternoon login patterns, where most are active, we can also see a peak in the evening between 7 PM and 8 PM. This specific time slot is interesting as the variations in cluster patterns started in the evening except for this particular time band. To understand any relationship between student demographics and chronotype clusters, we adopt the chi-square significant test and record the p-values that are listed in Table 6.3. From Table 6.3, we can observe a meaningful relationship between student ethnicity and chronotypes. Another significant relation is found between student enrollment type (full-time or part-time) and their chronotypes. This can be related to the varying course loads and a difference in the nonacademic workload of part-time students compared to full-time students. Finally, as a student progresses through their undergraduate years, we can see a shift in chronotypes.

Now that we understand different student chronotypes based on their login behaviors, the next step is to understand the importance of these chronotypes on student performance outcomes. To do this, we adopt a predictive modeling approach. In this approach, we use the student chronotype clusters as an input variable in addition to student login variables listed in Table 6.1. The improvements in prediction performance are decided by comparing the outcomes in this study with our earlier analysis [210] that is performed on the same students without chronotype variables. As mentioned in the methodology section, we adopt five machine learning models and perform cross-validation to evaluate these models based on R square and Root Mean Square Error (RMSE) values. Table 6.4 below shows the performance of predictive models with student chronotypes and login variables to predict the end of term GPA. The hyperparameters listed in Table 6.2 for these models are similar to our earlier study. From the results in Table 6.4, we can observe that the addition of the student chronotypes variable didn't have an impact on model performances.

Table 6.3: Student Demographics and their corresponding p-values based on chi-square test with Chronotype Clusters

Demographic	P-value
Gender	0.0571
Ethnicity	0.0001
Full Time or Part Time	0.007
Regular or Transfer	0.384
Student Year	0.018
GPA Bands	0.855

Table 6.4: Performance evaluation of predictive models based on R squared and RMSE values

Model	R Squared	RMSE
LR	0.259	0.733
DT	0.242	0.743
SVM	0.196	0.773
RF	0.363	0.683
GBT	0.360	0.680
Best Model [49]	0.362	0.622

6.4.2 Is there a causal relationship between student login behaviors, chronotypes, and their academic performance?

The second part of this study explores the causal relationships between student login variables, chronotypes, and performance outcomes. In this analysis, we first focus on identifying the relationship between all variables in this study. To do this, we apply a PC Variant algorithm in TETRAD that can handle both categorical and continuous variables. The composite variable input for the PC algorithm is calculated based on mCCA. In addition to the input mCCA composite features, we also input the knowledge component to control the directionality of cause and effect. In this case, we add Demographic variables as a cause as they cannot be varied. The login variables are both cause and effect. These variables act as a cause for end-of-term GPA but can be affected by student demographics. The feature weights related to mCCA are listed in Table D. 1. The causal relationships between variables are shown in Figure 6.5. From Figure 6.5, we can observe that the student academic performance has a direct cause and effect relationship with Normalized login volume and student prior performance. Our chronotype analysis showed a significant relationship between student chronotypes, student enrollment type (part-time or full-time), and student year. Other demographics like gender and type of admit stand independent from login variables or performance. This observation is also shown in the causal analysis, where we can see that their enrollment type and student year influence chronotypes. In addition to this, student regularity of logins affects chronotype clusters. Even though there are other causal relationships between different student variables, they are not directly impacting the performance. All the relationships displayed in Figure 6.5 are significant as we set a p-value cutoff at 0.05.

One of the significant drawbacks of the PC variant algorithm is the lack of confounder identification. To identify if the causal relations are more robust or if there is a presence of any confounder, we employ the GFCI algorithm in TETRAD. Additionally, we also explore the causal relationships by dividing student data into subsets based on their different demographic features. The interpretation of causal connectors related to GFCI is given in Table 6.5 below.

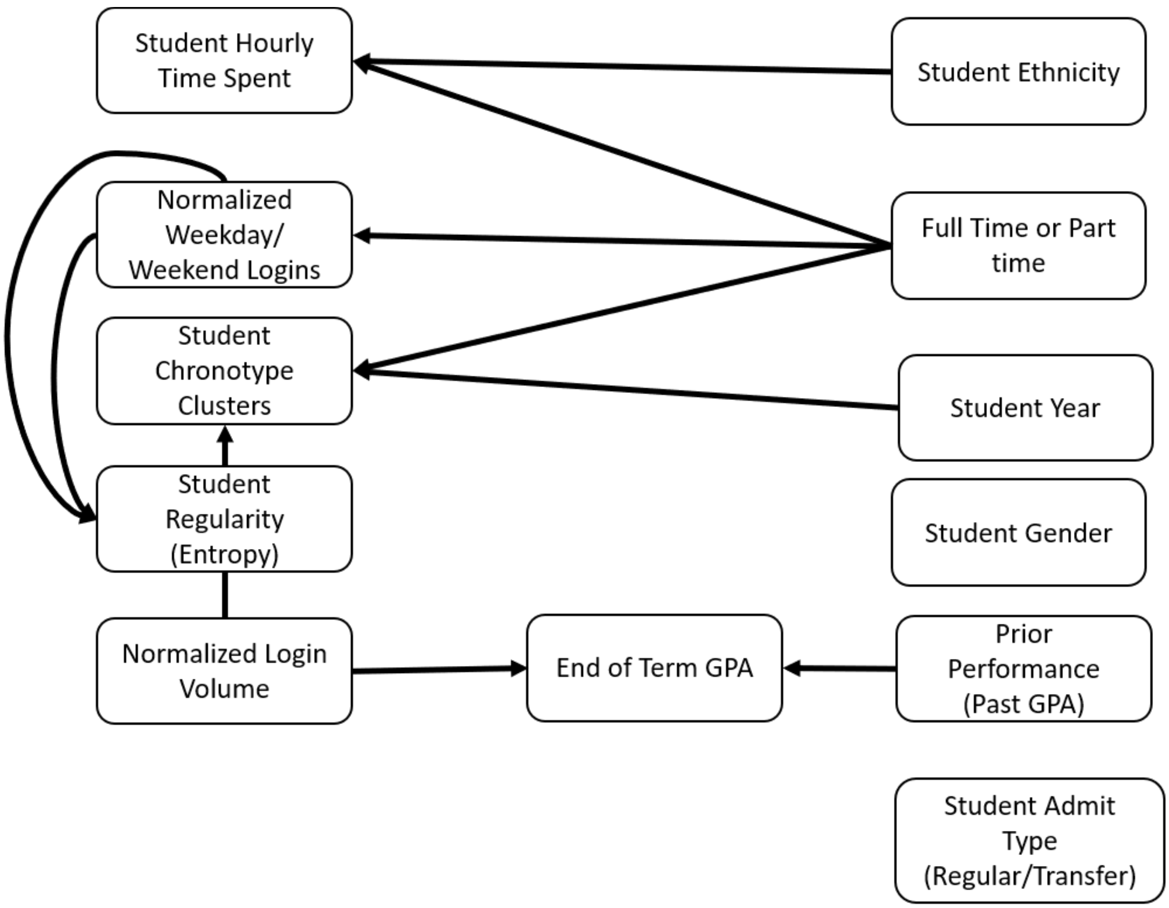


Figure 6.5: Causal relationship exploration between student demographics, login behaviors and chronotypes using PC Variant Algorithm.

Table 6.5: Graph connector type descriptions

Connector Type	Description
A $\longrightarrow$ B	A is a cause of B. It may be direct or indirect cause
A $\circ\!\!\longrightarrow$ B	Either A is a cause of B or there is an unmeasured confounder of A and B
A $\longleftrightarrow$ B	There is definitely an unmeasured confounder of A and B
A $\longrightarrow$ B	There is a direct relationship between A and B (no confounders)
A $\circ\!\!-\!\!\circ$ B	This can be any of the above cases or a combination of them. A relation exists but not clear on directionality of directness.

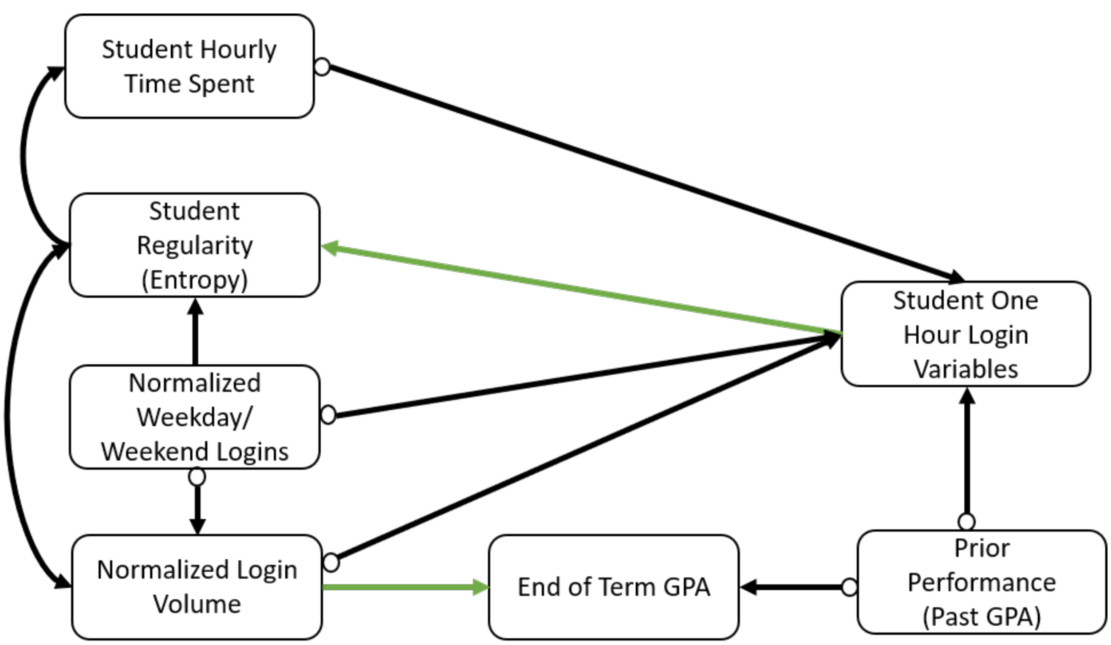


Figure 6.6: Causal relationship between Student login variables and their academic performance using GFCI algorithm

Figure 6.6 shows the output of the GFCI algorithm related to continuous student login variables. We can prove a direct relationship between student normalized login volume and their academic performance from this Figure 6.6. The relationship between student prior performance (Cumulative GPA till the start of the semester) and the current end of term GPA observed in PC variant analysis shown in Figure 6.5 also exists in GFCI output. The relationship between student prior performance (Cumulative GPA till the start of the semester) and the current end of term GPA observed in PC variant analysis shown in Figure 6.5 also exists in GFCI output. But it is not clear if this is a direct relationship if there might be an unmeasured confounder between these two. Consistent relations identified by PC Variant and GFCI methods are found between student login regularity & student login volume, student hourly login volume & regularity, and Weekend/WeekDay login volumes & login volumes as well as regularity. Even though multiple other relationships are common in both PC Variant analysis and GFCI for different variables, we can see no direct relationships that impact student performance except login volume and prior performance.

Figure 6.6 doesn’t fully specify causal relationships as there are some relationships with confounding factors. In addition to this, the GFCI causal graph in Figure 6.6 only shows the directionality of relationships but no quantifiable outcomes. The causal graph shown in Figure 6.6 is used to generate a Directed Acyclic Graph (DAG) based on domain knowledge. This DAG is used as an input to estimate linear Gaussian Structural Equation Models (SEM). We then calculate the model's goodness of fit and coefficients to study if the causal relationship between the features is as expected based on domain knowledge. For example, for the pair of variables prior performance and End of Term GPA, we believe that there is a direct relationship between these two as prior performance can only be a cause but not an effect. For other variables with confounding measures, we can expect any directionality of cause and effect relationship as they all are relevant and drawn from the similar statistic. This is the reason we assign a double-headed arrow to other relationships with confounding factors.

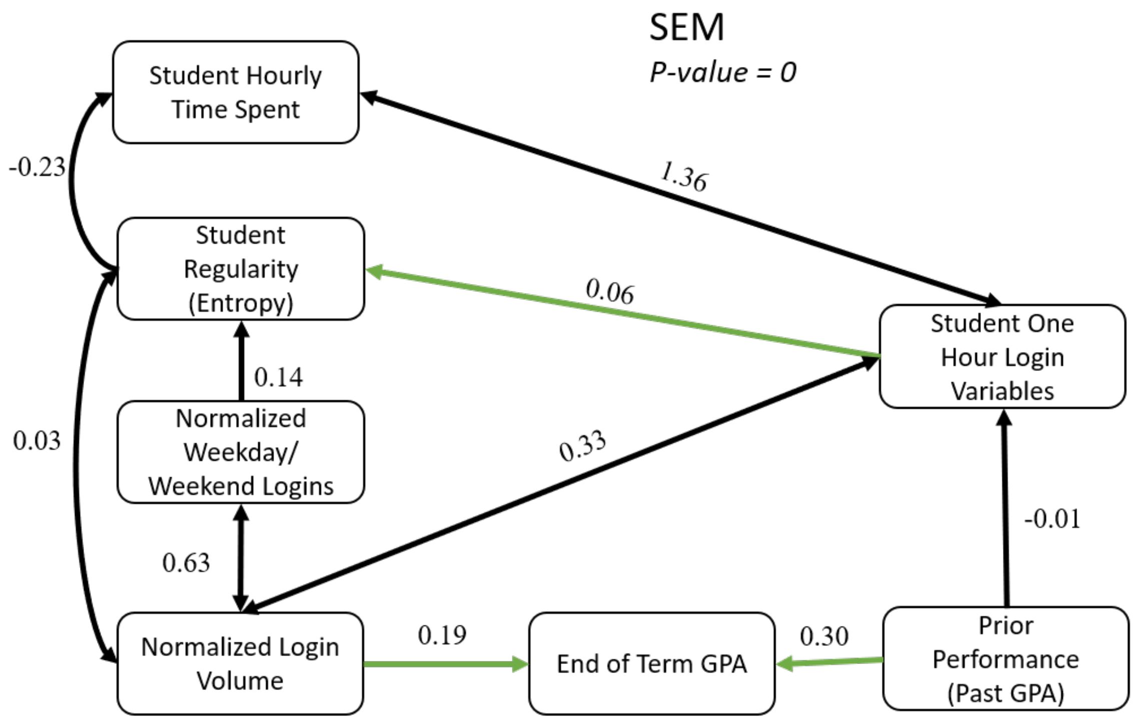
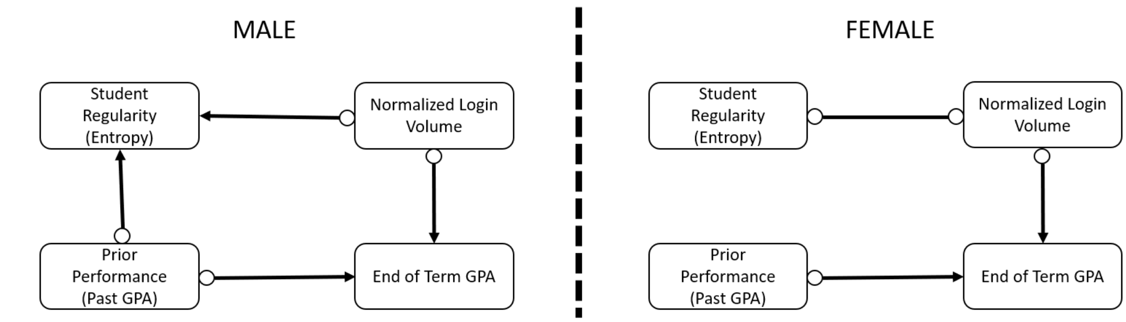


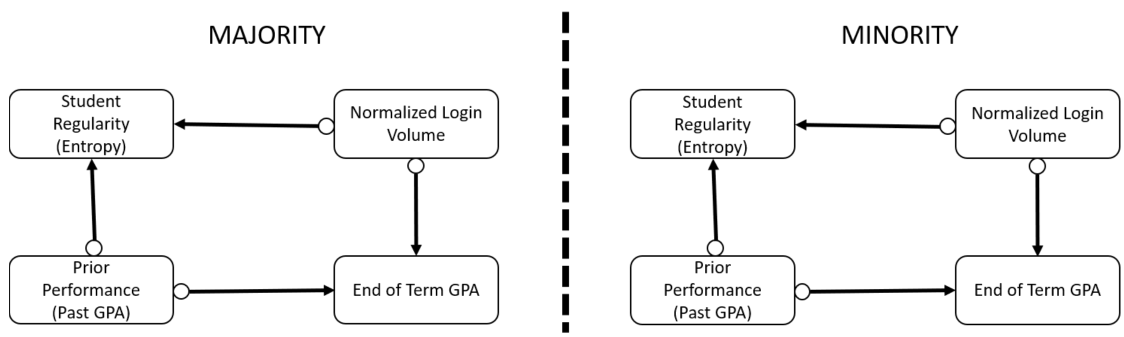
Figure 6.7: DAG derived from GFCI causal graph. Numerical values on directed arrows represent coefficients, and the values bidirectional arrows represent covariance reported by SEM.

Figure 6.7 represents DAG derived from the causal graph in Figure 6.6. The values indicated on arrows represent coefficients if it is a unidirectional arrow and covariance if it is a bidirectional arrow. From this DAG, we can observe that Login volume causes End of term GPA and if the login volume increases, student GPA increases as the coefficient value (0.19) is positive. A similar relationship can be observed between student prior performance and current term end GPA. The model's goodness of fit reveals a significant relationship between all the edges as all of their p values are close to 0 and less than 0.05.

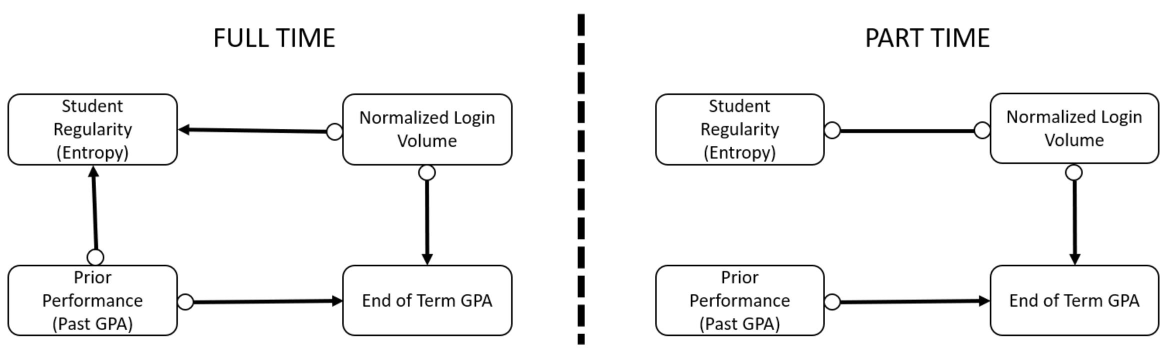
From Figure 6.7, we can confirm that the causal relationships observed in the GFCI algorithm are sensible.



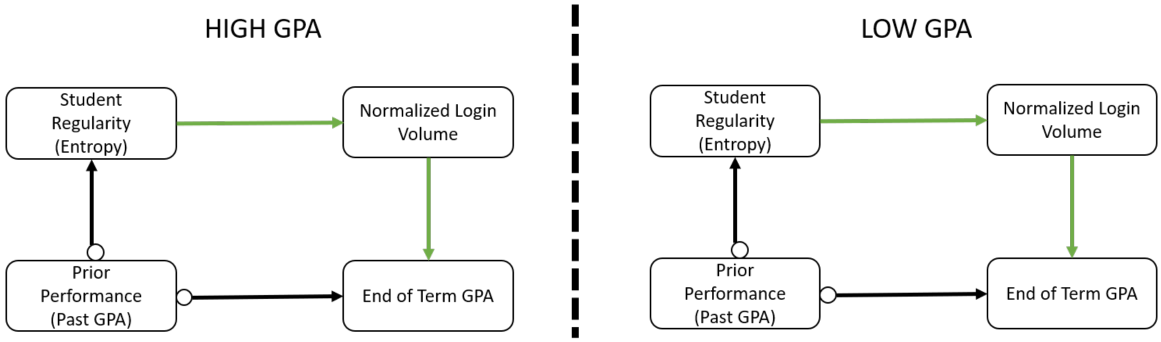
(a)



(b)



(c)



(d)

Figure 6.8: (a) Causal relationships based on gender-specific dataset (b) Causal relationships based on student ethnicity (c) Causal relationships based on student enrollment type (Full-time/Part-time) (d) Causal relationships based on student GPA

In the next step, we divide the data into subsets based on gender, race, and GPA to study causal relationships based on these demographics. We also reduce the number of variables based on the knowledge gained from overall causal relationships in PC Variant and GFCI approaches. The gender based feature weights related to mCCA are mentioned in tables Table D. 2 and Table D. 3. We also reduce the number of variables based on the knowledge gained from overall causal relationships. From Figure 6.8 (a), we can observe that there is no direct relationship between login variables and performance for different gender and race classifications. One difference between Male and Female gender is the impact of student prior performance on student login regularity.

There is a relationship between student regularity and their performance for Males, but for females, this does not exist. From Figure 6.8 (b), we do not see any significant difference in relationships between students from Majority races and students from Minority races. The mCCA feature weights related to ethnicity based datasets are shown in Table D. 4 and Table D. 5. One interesting observation is the causal relationship between login volume and performance for minority race students, which we observed in our earlier work [210] with predictive analytics as well. The relationship between prior performance and regularity exists in full-time students but not part-time students, as shown in Figure 6.8 (c). The corresponding mCCA weights for Full time and Part time students are shown in Table D. 6 and Table D. 7. There is a significant and direct relationship between login variables and student performance for students with varying bands of GPA (< 2 Low GPA & >3 High GPA), as shown in Figure 6.8 (d). The mCCA weights based on GPA are shown in Table D. 8 and Table D. 9. Even though there are no clear relationships without confounders based on gender and race, it is still valuable to learn that a relationship exists between student logins and their performance.

6.5 Discussion

With the increasing adoption of LMS systems in colleges and universities, there is growing interest in understanding student learning behaviors based on their interaction with these systems. Studies in Learning Analytics and Educational Data Mining focused on developing models that predict student performance based on their interaction data captured by these systems. The primary objective of most of these models is to predict student performance in the early stages of their academic career and develop intervention techniques that positively impact their academic performance. It is essential to study the relationship between student interaction variables and their performances to design these interventions efficiently. One of the most common methods is to explore relationships based on statistical significance tests. However, these tests don't specify the cause-and-effect relationship between student variables.

Causal inference and discovery methods were developed to mitigate this issue. We employ causal modeling methods to study the relationships between student login variables captured by LMS data and their academic performance as a step in this direction. In addition to this, we also study the impact of chronotypes explored in the biological field that related to human productivity. In addition to this, we also study the effect of chronotypes explored in the biological domain associated with human productivity on student academic performance.

Earlier studies showed that student self-regulation features encoded in LMS log data could be explored to develop effective interventions. From the analysis in this study, we observed that student login volume directly impacts student academic performance. The causal relationships revealed that student volume is also influenced by student self-regulation variables like student regularity in logins. This is an important find as this shows that triggering an intervention that increases student logins or their login regularity might improve their overall academic performance. The demographics-based study also showed the impact of logins is significant on students from different backgrounds. The relationships between self-regulation variables and performance are stronger in students at different GPA levels. These findings confirm that studies at the student level reveal valuable insights that are transferable for different demographics and can act as strong inputs for effective intervention development.

6.5.1 Key Contributions

One of the primary contributions of this study is the identification of student chronotypes based on their interaction variables. Earlier studies discussed the chronotype patterns in individuals and their impact on productivity [100][117][118]. These studies showed that individuals at younger ages are highly active in the morning, but this shifts to evening as they reach adolescence. The clustering methodology employed in this study showed a similar pattern where students in an undergraduate university showed high activity during the afternoon to evening hours. Prior studies also showed that these chronotypes have statistically significant relationships with academic achievement. However, these studies are primarily performed on a single domain or course [218]. Our analysis explicitly targets students as a single entity and aggregates their activity across all enrolled courses.

The statistical findings in our work showed significant relationships between student demographics and chronotypes but not between chronotypes and their performance. In addition to this, we also performed a predictive modeling methodology to study if these chronotypes contribute to student performance prediction. The findings from this analysis revealed that student performance prediction stayed similar with and without the chronotype variables. These observations align with some earlier studies [117][218] that contradicted the finds of relationships between chronotypes and performance. Overall, research on the relationship between student performance and chronotypes needs more investigation as the findings between different studies are inconsistent.

In addition to studying student chronotypes, we also analyzed the causal relationships between student login variables, chronotypes, and performances. The findings from this analysis showed that student login volume and prior performance have a direct cause-and-effect relationship with academic performance. This finding validated our earlier work [49] that showed student login variables played a significant role in predicting the end of term GPA. Further analysis of causal relationships based on demographics revealed that login variables related to low-performance students also showed a conclusive effect on student performance. This finding also validates our earlier work related to student-centric models that emphasized the contribution of logins on low GPA student performance prediction. Even though our earlier study showed a significant impact of logins on GPA prediction based on ethnicity, the causal analysis didn't reveal any conclusive results. The findings showed that there is a cause-and-effect relationship, but it might be due to confounding factors.

There are also some significant limitations in this study. The data captured by LMS is only a snapshot of activity in students' day-to-day lives. Student performance factors can be influenced by many other external factors like study environment, family background, and student perception towards a course. The login-based chronotypes discussed in this study are only a part of time management. Students' time management can be observed from multiple other factors like assignment submissions, time spent on course contents, and exploring time lag between lecture delivery and student reading content access. Another limitation is related to the dataset. This aggregate-level dataset only captures the student login information but not content-level information. Content level information is much more fine-grained and provides much more insights into what a student might be working on when they are logging into the system. One significant challenge with content-level information is their diversity based on course and instructor style. This will make it hard to extract student-level aggregated features.

To conclude, the causal analysis in this study strengthened our earlier findings that showed significant relationships between student login variables and their performance. The conclusions of this study are valuable to Educational Data Mining and Learning

Analytics domains as they support the design and development of interventions techniques based on LMS variables to improve student performance. As the data in this study is collected till the middle of the semester, the development and deployment of interventions at this stage provides valuable time to students for improvement and contribute to their academic achievement.

Chapter 7: Conclusion & Future Work

This research explored student learning behaviors based on academic emotions and self-regulation characteristics like regularity and chronotypes. The findings from this work are detailed below.

1. Academic Emotions: Tracking student academic emotions help researchers to develop methods that support emotional interventions that benefit student learning. As part of this work, we proposed a novel method to track student academic emotions based on changes rather than specific states. These computational models developed as part of this study to predict academic emotions from LMS systems proved efficient compared to earlier works.
2. Academic Emotions & Student Outcomes: One of the critical aspects of this research is to study if the developed models support student outcomes. We adopted computational models and developed a voting machine system to predict student career choices to explore this. The experimental results of the proposed models showed higher performance compared to earlier research published in this domain. In addition to this, we also explore the relationship between student interaction data captured by intelligent tutoring systems and emotions with student outcomes. We developed a novel explanation method based on concepts in explainable AI. This method showed that positive academic emotions like concentration supported academic outcomes and academic emotions like boredom and frustration negatively impacted student outcomes.
3. Student-Centered Modeling: The second part of this work focused on modeling students based on their interactions with LMS. As part of this work, we developed

models that utilize aggregate level student LMS login data to predict their overall academic performance. This is a novel approach compared to existing research that focused on course-level interaction data, supporting more efficient intervention designs. Moreover, we observed a strong relationship between student activity of the LMS system with their academic performance. The demographic-based studies revealed a strong correlation between student login volume and their outcome for students with low GPA and students from minority races respectively. The insights provided by this work prompt new questions to investigate the reasons behind the impact of login volumes on a student with a lower GPA and students from a minority race as the login data from LMS is a snapshot of student activity and can be influenced by many other external factors like economic conditions, system design, and technology availability. In addition to this, the findings from this work also support promoting LMS usage by instructors as it helps predict student performance early in the semester.

4. Chronotypes and Causality: Student self-regulation strategies have a significant impact on their learning outcomes. As part of this work, we explored a segment of self-regulation that focuses on time characteristics referred to as chronotypes. Student hourly login data is captured from the LMS system to develop clustering models that show some common time patterns among students. The outcomes demonstrated that most of the students are highly active during the evening. This finding is in line with earlier studies in chronobiology that reported human chronotypes to shift from morningness to eveningness by the age of 20. Even though earlier course-specific studies showed some relation between student

chronotypes and academic performance, our predictive analytics study finds no impact of student chronotypes on performance. In addition to this chronotype study, we also explored the causal relationships between student LMS login behaviors and their academic performance as understanding this will improve intervention designs. We observed a strong causal relationship between student login volume and academic performance from our experiments. We also observed a direct relationship between student's prior performance and their current performance. The causal relationships observed in this study support the development of intervention methods that helps modulate student logins to improve their performance. These intervention methods can be developed at a system level to be effective irrespective of the student cohort. One such example is currently implemented at UMBC. The blackboard at UMBC displays the number of logins and compares them with the average logins in that class. However, it is still important to educate students on the importance of the login indicators that contribute to their academic performance. Additional qualitative work is also needed to study what factors influence students to log in frequently or less so that the development of interventions can be much more efficient.

Even though the findings from this research are important and valuable to learning analytics and the education domain, there is much room for improvement that can be explored in future studies. Some directions for future work in this area are detailed below.

1. Emotion-specific interventions: The primary purpose of most predictive models is to develop intervention techniques that improve current results. Even though our study and prior studies in academic emotions showed the solid predictive power of

computational models in affect detection, it is still unclear how these predictions can be used to design efficient interventions. One important direction is to study how the student interaction with LMS systems can explain these academic emotions. This understanding will help develop targeted interventions to modulate or amplify specific emotions that support academic outcomes or learning behavior. In addition to this, it is also important to integrate more concepts from psychology with educational data mining to develop sustainable and explainable models.

2. Content Level Analysis: One of the essential aspects of student-centric modeling is to generate aggregate-level student interaction features across all courses. Our research is the first step to aggregate student login-related features across all courses a student enrolled. However, our work didn't utilize any content-level features. Content level features give a deeper understanding of student behaviors based on the material they access and spend time on. One challenge is to aggregate content level features as they vary from one course to another based on instructor style and course type. More research in this direction is needed to define aggregate content level features to improve student-centric models' prediction performance and generate more insights at an individual level.
3. Ubiquitous Learning System: Most research in learning analytics and educational data mining focuses on understanding student learning behaviors to support their development and help them on the way to become lifelong learners. It is essential to focus on developing ubiquitous learning systems that consider different factors related to student learning like emotions, behaviors, learning patterns, and skills to deliver content efficiently and effectively. These systems should have the

capabilities to predict factors that influence student learning in real-time and provide personalized interventions to reach their academic goals.

Appendix A

Table A. 1: Cross validation performance metrics of conventional models on 2-clip datasets without dimensionality reduction

Model	Concentration - Boredom			Concentration - Frustration			Concentration - Confusion			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
DT	0.676	0.361	0.479	0.5	0.046	0.433	0.5	0.014	0.430	0.598	0.215	0.481
RF	0.817	0.459	0.431	0.637	-0.019	0.405	0.652	-0.007	0.399	0.801	0.411	0.431
SVM	0.823	0.509	0.416	0.667	0.091	0.402	0.631	-0.007	0.410	0.823	0.484	0.409
LR	0.497	0.116	0.646	0.384	-0.052	0.704	0.428	-0.020	0.699	0.497	0.203	0.621
AMLP	0.789	0.472	0.473	0.667	0.091	0.419	0.683	0.004	0.431	0.808	0.439	0.480
NN	0.796	0.441	0.474	0.595	0.135	0.490	0.734	0.144	0.454	0.761	0.407	0.495

Table A. 2: Cross validation performance metrics of conventional models on 2-clip datasets with dimensionality reduction

Model	Concentration - Boredom			Concentration - Frustration			Concentration - Confusion			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
DT	0.668	0.37	0.48	0.5	0.127	0.402	0.5	0.014	0.43	0.632	0.27	0.471
RF	0.829	0.517	0.427	0.764	0.144	0.379	0.755	0.085	0.387	0.814	.387	0.428
SVM	0.838	0.585	0.4	0.786	0.107	0.382	0.715	0.023	0.395	0.834	0.483	0.406
LR	0.654	0.339	0.546	0.525	0.081	0.625	0.512	0.071	0.620	0.650	0.330	0.538
AMLP	0.816	0.517	0.464	0.731	0.263	0.424	0.713	0.131	0.427	0.805	0.444	0.478
NN	0.789	0.446	0.475	0.739	0.296	0.453	0.783	0.31	0.416	0.810	0.449	0.466

Table A. 3: Cross validation performance metrics of deep learning models on 2-clip datasets without dimensionality reduction

Model	Concentration - Boredom			Concentration - Frustration			Concentration - Confusion			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
CNN	0.618	0.236	0.632	0.5	0	0.454	0.512	0.038	0.454	0.561	0.126	0.621
SimpleRNN	0.664	0.328	0.577	0.561	0.126	0.519	0.582	0.152	0.525	0.637	0.271	0.599
LSTM	0.678	0.356	0.562	0.545	0.077	0.525	0.590	0.181	0.522	0.683	0.383	0.535
GRU	0.667	0.333	0.577	0.524	0.068	0.531	0.579	0.152	0.540	0.683	0.372	0.547

Table A. 4: Cross validation performance metrics of deep learning models on 2-clip datasets with dimensionality reduction

Model	Concentration - Boredom			Concentration - Frustration			Concentration - Confusion			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
CNN	0.730	0.466	0.540	0.5	0	0.454	0.496	-0.004	0.481	0.648	0.303	0.553
SimpleRNN	0.723	0.441	0.527	0.612	0.212	0.498	0.666	0.331	0.461	0.731	0.459	0.513
LSTM	0.73	0.454	0.524	0.566	0.124	0.507	0.627	0.245	0.491	0.693	0.399	0.532
GRU	0.657	0.309	0.588	0.589	0.197	0.484	0.586	0.181	0.497	0.663	0.331	0.565

Appendix B

Table B. 1: Cross validation performance metrics of conventional models on 3-clip datasets without dimensionality reduction

Model	Conc – Boredom - Conc			Conc – Frustration - Conc			Conc – Confusion - Conc			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
DT	0.500	0.269	0.471	0.500	-0.031	0.436	0.500	0.027	0.420	0.615	0.257	0.486
RF	0.791	0.297	0.428	0.701	0.006	0.373	0.699	-0.006	0.368	0.828	0.487	0.432
SVM	0.807	0.495	0.409	0.715	0.087	0.383	0.650	-0.027	0.394	0.837	0.479	0.411
LR	0.411	-0.023	0.715	0.403	0.016	0.695	0.414	0.003	0.696	0.388	0.034	0.695
AMLP	0.676	0.245	0.485	0.778	0.021	0.406	0.539	0.015	0.395	0.788	0.268	0.555
NN	0.819	0.485	0.446	0.691	0.220	0.419	0.643	0.104	0.464	0.782	0.420	0.498

Table B. 2: Cross validation performance metrics of conventional models on 3-clip datasets with dimensionality reduction

Model	Conc – Boredom - Conc			Conc – Frustration - Conc			Conc – Confusion - Conc			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
DT	0.5	0.106	0.481	0.500	0.111	0.394	0.500	0.057	0.388	0.647	0.372	0.470
RF	0.736	0.173	0.449	0.720	0.146	0.370	0.732	0.013	0.365	0.839	0.501	0.415
SVM	0.751	0.304	0.435	0.672	0.044	0.383	0.716	-0.018	0.384	0.834	0.494	0.409
LR	0.485	0.093	0.665	0.517	0.046	0.633	0.603	0.137	0.578	0.571	0.219	0.618
AMLP	0.758	0.358	0.473	0.778	0.143	0.385	0.685	0.174	0.368	0.848	0.49	0.434
NN	0.748	0.402	0.463	0.701	0.168	0.444	0.753	0.228	0.405	0.776	0.392	0.499

Table B. 3: Cross validation performance metrics of deep learning models on 3-clip datasets without dimensionality reduction

Model	Conc – Boredom - Conc			Conc – Frustration - Conc			Conc – Confusion - Conc			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
CNN	0.59	0.199	0.542	0.498	-0.006	0.425	0.5	0	0.418	0.582	0.165	0.651
SimpleRNN	0.663	0.331	0.540	0.575	0.159	0.472	0.535	0.091	0.497	0.630	0.261	0.606
LSTM	0.643	0.281	0.561	0.662	0.321	0.583	0.515	0.052	0.526	0.540	0.087	0.509
GRU	0.616	0.233	0.584	0.483	-0.018	0.522	0.543	0.054	0.526	0.643	0.278	0.599

Table B. 4: Cross validation performance metrics of deep learning models on 3-clip datasets with dimensionality reduction

Model	Conc – Boredom - Conc			Conc – Frustration - Conc			Conc – Confusion - Conc			Always Concentration		
	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE	AUC	Kappa	RMSE
CNN	0.512	0.022	0.600	0.512	0.033	0.417	0.534	0.084	0.437	0.653	0.306	0.607
SimpleRNN	0.619	0.239	0.584	0.538	0.093	0.496	0.587	0.178	0.463	0.686	0.371	0.559
LSTM	0.632	0.262	0.573	0.536	0.066	0.490	0.566	0.145	0.473	0.694	0.387	0.549
GRU	0.562	0.128	0.617	0.522	0.042	0.490	0.604	0.196	0.480	0.677	0.351	0.571

Appendix C

Table C. 1: Attributes with high impact on predictions

Attribute	Correlation (%)	ID-ness (%)	Stability (%)	Category
NumActions	0.39	0.15	0.84	H
timeTaken	0.03	NA	3.62	H
correct	0.18	0	59.59	H
original	0.15	0	71.26	H
hint	0.26	0	70.05	H
hintCount	0.29	0.01	55.54	H
hintTotal	0.31	0.01	49.13	H
attemptCount	0.11	0.02	48.73	H
frPast5HelpRequest	0.15	0	31.17	H
frPast8HelpRequest	0.13	0	29.82	H
past8BottomOut	0.07	0	86.56	H
totalFrPastWrongCount	0.01	0.02	33.11	H
timeSinceSkill	0.03	4.36	71.02	H
totalFrAttempted	0.01	0.36	0.46	H
totalFrSkillOpportunities	0.03	0.06	13.39	H
endsWithScaffolding	0.08	0	61.64	H
frTotalSkillOpportunitiesScaffolding	0.09	0.03	34.45	H
frIsHelpRequestScaffolding	0.14	0	64.65	H
sumRight	0.04	0.27	0.56	H
timeGreater10SecAndNextActionRight	0.34	0	75.53	H
sumTimePerSkill	0.02	NA	0.21	H
totalTimeByPercentCorrectForskill	0.14	NA	14.1	H
manywrong	0.54	0	66.33	H
RES_GAMING	0.19	NA	15.48	H
Ln-1	0.57	NA	0.52	H
Ln	0.67	NA	0.5	H

Table C. 2: Attributes with Medium impact on predictions

Attribute	Correlation (%)	ID-ness (%)	Stability (%)	Category
skill_ID	0	0.03	8.3	M
scaffold	0	0	60.74	M
frIsHelpRequest	0	0	72.54	M
totalFrPercentPastWrong	0	NA	37.78	M
frPast5WrongCount	0	0	47.3	M
frPast8WrongCount	0	0	41.74	M
totalFrTimeOnSkill	0	1.29	14	M
totalFrSkillOpportunitiesByScaffolding	0.01	NA	44.48	M
sumTimePerSkill	0	NA	44.48	M
RES_BORED	0	NA	67.59	M
RES_CONFUSED	0	NA	80.44	M
RES_FRUSTRATED	0	NA	59.46	M
RES_OFFTASK	0.01	NA	58.09	M

Table C. 3: Attributes with Low impact on predictions

Attribute	Correlation (%)	ID-ness (%)	Stability (%)	Category
bottomHint	0.04	0	94.54	L
stlHintUsed	0	0	99.65	L
frWorkingInSchool	0	0	97.2	L
responseIsFillln	0.02	0	97.9	L
responseIsChosen	?	0	100	L
endsWithAutoScaffolding	0.01	0	99.45	L
frTimeTakenOnScaffolding	0	NA	28.77	L
timeGreater5Secprev2Wrong	0.01	0	95.55	L
helpAccessUnder2Sec	0.06	0	95.19	L
consecutiveErrorsInRow	0	0.02	92.63	L
Prev5count	0	0	99.09	L
timeOver80	0.03	0	91.02	L
RES_CONCENTRATING	0	NA	4.45	L

Appendix D

Table D. 1: Composite variables with associated original features and weights estimated by mCCA for all students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Max Volume	0.36
	Median Volume	0.93
Login Regularity (Entropy)	Mean Entropy	-0.36
	Median Entropy	-0.93
Hourly Login Volumes	H1 to H9	< 0.18
	H10 to H24	0.24
Hourly Time Spent	H1 to H8	0.14 to 0.17
	H9 to H21	0.21 to 0.26
	H22 to H24	0.19
Weekday/Weekend Login Volumes	WeekDay	0.99
	WeekEnd	0.1

Table D. 2: Composite variables with associated original features and weights estimated by mCCA for MALE students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Min Volume	-0.3
	Max Volume	-0.52
	Mean Volume	-0.54
	Median Volume	-0.55
	Standard Deviation Volume	-0.2
Login Regularity (Entropy)	Min Entropy	0.41
	Max Entropy	0.43
	Mean Entropy	0.57
	Median Entropy	0.57
	Standard Deviation Entropy	0.06
	Skewness Entropy	-0.02
	Kurtosis Entropy	0.002

Table D. 3: Composite variables with associated original features and weights estimated by mCCA for FEMALE students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Mean	0.11
	Median	0.99
Login Regularity (Entropy)	Mean	-0.99
	Median	-0.11

Table D. 4: Composite variables with associated original features and weights estimated by mCCA for Majority Race (White & Asian) students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Mean	-0.11
	Median	-0.99
Login Regularity (Entropy)	Mean	0.99
	Median	0.11

Table D. 5: Composite variables with associated original features and weights estimated by mCCA for Minority Race students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Mean	0.11
	Median	0.99
Login Regularity (Entropy)	Mean	-0.11
	Median	-0.99

Table D. 6: Composite variables with associated original features and weights estimated by mCCA for Full-time students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Mean	-0.56
	Median	-0.59
	Min	-0.13
	Max	-0.53
	Std	-0.07
Login Regularity (Entropy)	Mean	0.62
	Median	0.62
	Min	0.30
	Max	0.37

Table D. 7: Composite variables with associated original features and weights estimated by mCCA for Part-time students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Max	-0.99
	Median	-0.11
Login Regularity (Entropy)	Mean	0.99
	Median	0.11

Table D. 8: Composite variables with associated original features and weights estimated by mCCA for Low GPA (<2) students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Mean	0.11
	Median	0.99
Login Regularity (Entropy)	Mean	-0.11
	Median	-0.99

Table D. 9: Composite variables with associated original features and weights estimated by mCCA for High GPA (>=3) students

Composite Variable	Feature Name	Weights
Normalized Login Volume	Mean	-0.36
	Median	-0.93
Login Regularity (Entropy)	Mean	0.93
	Median	0.36

Bibliography

[1] Isik, U., Tahir, O. E., Meeter, M., Heymans, M. W., Jansma, E. P., Croiset, G., & Kusurkar, R. A. (2018). Factors Influencing Academic Motivation of Ethnic Minority Students: A Review. SAGE Open. https://doi.org/10.1177/2158244018785412

[2] Desforges, C., & Abouchaar, A. (2003). The impact of parental involvement, parental support and family education on pupil achievement and adjustment: A literature review (Vol. 433). London: DfES.

[3] Farkas, G. (2003). Racial disparities and discrimination in education: What do we know, how do we know it, and what do we need to know?. Teachers College Record, 105(6), 1119-1146.

[4] Villavicencio, F. T., & Bernardo, A. B. (2013). Positive academic emotions moderate the relationship between self-regulation and academic achievement. British Journal of Educational Psychology, 83(2), 329-340.

[5] Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., ... & Gabrieli, J. D. (2014). Cognitive skills, student achievement tests, and schools. Psychological Science, 25(3), 736-744.

[6] Ghaith, G. M. (2002). The relationship between cooperative learning, perception of social support, and academic achievement. System, 30(3), 263-273.

[7] Dabbagh, N., & Kitsantas, A. (2013). Using learning management systems as metacognitive tools to support self-regulation in higher education contexts. In International handbook of metacognition and learning technologies (pp. 197-211). Springer, New York, NY.

[8] Colorado, J. T., & Eberle, J. (2012). Student demographics and success in online learning environments.

[9] Tempelaar, D. (2017). How Dispositional Learning Analytics Helps Understanding the Worked-Example Principle. International Association for Development of the Information Society.

[10] Pekrun, R. (2011). Emotions as drivers of learning and cognitive development. In New perspectives on affect and learning technologies (pp. 23-39). Springer, New York, NY.

[11] Pekrun, R. (2016). Academic emotions. Handbook of motivation at school, 2, 120-144.

[12] Spangler, G., Pekrun, R., Kramer, K., & Hofmann, H. (2002). Students' emotions, physiological reactions, and coping in academic exams. Anxiety, Stress & Coping, 15(4), 413-432.

[13] Zeidner, M. (1995). Coping with examination stress: Resources, strategies, outcomes. Anxiety, stress, and coping, 8(4), 279-298.

[14] Zeidner, M. (2007). Test anxiety in educational contexts: What I have learned so far. Emotion in education, 165184.

[15] Zeidner, M. (1998). Test anxiety: The state of the art. Springer Science & Business Media.

[16] Efklides, A. (Ed.). (2005). Feelings and emotions in the learning process. Elsevier.

[17] Linnenbrink, E. A. (2006). Emotion research in education: Theoretical and methodological perspectives on the integration of affect, motivation, and cognition. Educational psychology review, 18(4), 307-314.

[18] Linnenbrink, E. A. (2007). The role of affect in student learning: A multi-dimensional approach to considering the interaction of affect, motivation, and engagement. In Emotion in education (pp. 107-124). Academic Press.

[19] Hussain, M. S., AlZoubi, O., Calvo, R. A., & D'Mello, S. K. (2011, June). Affect detection from multichannel physiology during learning sessions with AutoTutor. In International Conference on Artificial Intelligence in Education (pp. 131-138). Springer, Berlin, Heidelberg.

[20] Nye, B. D., Karumbaiah, S., Tokel, S. T., Core, M. G., Stratou, G., Auerbach, D., & Georgila, K. (2018, June). Engaging with the Scenario: Affect and Facial Patterns from a Scenario-Based Intelligent Tutoring System. In International Conference on Artificial Intelligence in Education (pp. 352-366). Springer, Cham.

[21] Bosch, N., D'mello, S. K., Ocumpaugh, J., Baker, R. S., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. ACM Transactions on Interactive Intelligent Systems (TiiS), 6(2), 1-26.

[22] Paquette, L., Rowe, J., Baker, R., Mott, B., Lester, J., DeFalco, J., ... & Georgoulas, V. (2016). Sensor-Free or Sensor-Full: A Comparison of Data Modalities in Multi-Channel Affect Detection. International Educational Data Mining Society.

[23] Hijazi, S. T., & Naqvi, S. M. M. (2006). FACTORS AFFECTING STUDENTS'PERFORMANCE. Bangladesh e-journal of Sociology, 3(1).

[24] Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017, June). Improving sensor-free affect detection using deep learning. In International Conference on Artificial Intelligence in Education (pp. 40-51). Springer, Cham.

[25] Zimmerman, B. J. (2013). Theories of self-regulated learning and academic achievement: An overview and analysis. In Self-regulated learning and academic achievement (pp. 10-45). Routledge.

[26] Schunk, D. H. (2012). Social cognitive theory.

[27] Miltiadou, M., & Savenye, W. C. (2003). Applying social cognitive constructs of motivation to enhance student success in online distance education. AACE journal, 11(1), 78-95.

[28] Childers, C., Williams, K., & Kemp, E. (2014). Emotions in the classroom: Examining environmental factors and student satisfaction. Journal of Education for Business, 89(1), 7-12.

[29] Cleveland-Innes, M., & Campbell, P. (2012). Emotional presence, learning, and the online learning environment. The International Review of Research in Open and Distributed Learning, 13(4), 269-292. 26

[30] Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: the impact of agent emotion and gender. Journal of Computer Assisted Learning, 23(3), 220-234.

[31] Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child development, 82(1), 405-432.

[32] Hazari, Z., Sadler, P. M., & Sonnert, G. (2013). The science identity of college students: Exploring the intersection of gender, race, and ethnicity. Journal of College Science Teaching, 42(5), 82-91.

[33] Goel, S. (2011). An overview of selected theories about student learning. Online Submission.

[34] Sinatra, G. M., Broughton, S. H., & Lombardi, D. O. U. G. (2014). Emotions in science education. International handbook of emotions in education, 415-436.

[35] Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. Child development, 82(1), 405-432.

[36] Valdesolo, P., Shtulman, A., & Baron, A. S. (2017). Science is awe-some: The emotional antecedents of science learning. Emotion Review, 9(3), 215-221.

[37] Spicer, J. (1997). Systems analysis of stress and coping: A testing proposition. Journal of health psychology, 2(2), 167-170.

[38] Churchill, D. (2017). Emerging possibilities for design of digital Resources for learning. In Digital Resources for Learning (pp. 227-246). Springer, Singapore.

[39] Lazarus, R. S. (2006). Stress and emotion: A new synthesis. Springer Publishing Company.

[40] Turner, F. D., & Smith, J. K. (2015). A comparative study on the stress levels of Black, White, Asian, and Latino undergraduate students. Journal of Research Initiatives, 1(3), 6.

[41] Karplus, S. S. (2017). Integrating academic library resources and learning management systems: The library Blackboard site. Education Libraries, 29(1), 5-11.

[42] Mouakket, S., & Bettayeb, A. M. (2015). Investigating the factors influencing continuance usage intention of Learning management systems by university instructors. International Journal of Web Information Systems.

[43] Zheng, Y., Wang, J., Doll, W., Deng, X., & Williams, M. (2018). The impact of organisational support, technical support, and self-efficacy on faculty perceived benefits of using learning management system. Behaviour & Information Technology, 37(4), 311-319.

[44] Kim, J., & Do, J. (2016). Learning management system: Medium for Interactive communication. International Journal of Applied Engineering Research, 11(2), 1073-1076.

[45] Waheed, M., Kaur, K., Ain, N., & Hussain, N. (2016). Perceived learning outcomes from Moodle: An empirical study of intrinsic and extrinsic motivating factors. Information Development, 32(4), 1001-1013.

[46] Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. Computers & Education, 98, 157-168.

[47] Kovanović, V., Joksimović, S., Gašević, D., Owers, J., Scott, A. M., & Woodgate, A. (2016, April). Profiling MOOC course returners: how does student behavior change between two course enrollments?. In Proceedings of the third (2016) ACM conference on learning@ scale (pp. 269-272).

[48] Somerfield, M. R., & McCrae, R. R. (2000). Stress and coping research: Methodological challenges, theoretical advances, and clinical applications. American psychologist, 55(6), 620.

[49] Reich, J. (2015). Rebooting MOOC research. Science, 347(6217), 34-35.

[50] Timms, M. J. (2016). Letting artificial intelligence in education out of the box: educational cobots and smart classrooms. International Journal of Artificial Intelligence in Education, 26(2), 701-712.

[51] Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. International Journal of Artificial Intelligence in Education, 26(2), 582-599.

[52] André, E., Baker, R., Hu, X., Rodrigo, M. T., & du Boulay, B. (2017). Artificial Intelligence in Education. In proceedings of the18th International Conference on AIED, Wuhan, China.

[53] Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). Contemporary educational psychology, 36(1), 36-48.

[54] Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. Educational psychologist, 37(2), 91-105.

[55] Torres, V., Reiser, A., LePeau, L., Davis, L., & Ruder, J. (2006). A model of first-generation Latino/a college students' approach to seeking academic information. NACADA journal, 26(2), 65-70.

[56] Thomas, L. J., Parsons, M., & Whitcombe, D. (2019). Assessment in Smart Learning Environments: Psychological factors affecting perceived learning. Computers in Human Behavior, 95, 197-207.

[57] Zhang, J., Gong, J., & Barnes, L. (2017, July). HCNN: heterogeneous convolutional neural networks for comorbid risk prediction with electronic health records. In 2017 IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE) (pp. 214-221). IEEE.

[58] Wolf, L., Stidham, A. W., & Ross, R. (2015). Predictors of stress and coping strategies of US accelerated vs. generic baccalaureate nursing students: An embedded mixed methods study. Nurse Education Today, 35(1), 201-205.

[59] Labrague, L. J., McEnroe-Petitte, D. M., Gloe, D., Thomas, L., Papathanasiou, I. V., & Tsaras, K. (2017). A literature review on stress and coping strategies in nursing students. Journal of Mental Health, 26(5), 471-480.

[60] Chang, J. (2018, October). Influencing Factors and Coping Strategies of Students' Engagement in the Context of Internet Education. In 2018 9th International

Conference on Information Technology in Medicine and Education (ITME) (pp. 517-520). IEEE.

[61] Al Shaibah, A., Condra, M., Jama, S., & Stewart, M. (2017). Q Success: Supporting underrepresented students in transition. Higher Education Quality Council of Ontario.

[62] Murray, M. A. (2016). Identity compatibility, career adaptability, and adaptive coping as predictors of college women's commitment in STEM majors. State University of New York at Albany.

[63] Williams, S. (2017). Beyond access towards success for first-generation college students of underrepresented ethnic backgrounds: The role of college adjustment and perceived stressors on academic achievement during the first year.

[64] Hascher, T. (2010). Learning and Emotion: perspectives for theory and research. European Educational Research Journal, 9(1), 13-28.

[65] Espinosa, L. (2011). Pipelines and pathways: Women of color in undergraduate STEM majors and the college experiences that contribute to persistence. Harvard Educational Review, 81(2), 209-241.

[66] Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. International Journal of Human-Computer Studies, 68(4), 223-241.

[67] Masters, J. C., Barden, R. C., & Ford, M. E. (1979). Affective states, expressive behavior, and learning in children. Journal of Personality and Social Psychology, 37(3), 380.

[68] D'Mello, S. (2013). A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. Journal of Educational Psychology, 105(4), 1082.

[69] Ocumpaugh, J., San Pedro, M. O., Lai, H. Y., Baker, R. S., & Borgen, F. (2016). Middle school engagement with mathematics software and later interest and self-efficacy for STEM careers. Journal of Science Education and Technology, 25(6), 877-887.

[70] Yeung, C. K., Lin, Z., Yang, K., & Yeung, D. Y. (2018). Incorporating Features Learned by an Enhanced Deep Knowledge Tracing Model for STEM/Non-STEM Job Prediction. arXiv preprint arXiv:1806.03256.

[71] Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017, June). Improving sensor-free affect detection using deep learning. In International conference on artificial intelligence in education (pp. 40-51). Springer, Cham.

[72] Wang, Y., Heffernan, N. T., & Heffernan, C. (2015, March). Towards better affect detectors: effect of missing skills, class features and common wrong answers. In Proceedings of the Fifth International Conference on Learning Analytics and Knowledge (pp. 31-35).

[73] Shum, S. B., & Crick, R. D. (2012, April). Learning dispositions and transferable competencies: pedagogy, modelling and learning analytics. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 92-101).

[74] Shum, S. B., Ferguson, R., & Martinez-Maldonado, R. (2019). Human-centred learning analytics. Journal of Learning Analytics, 6(2), 1-9.

[75] Dawson, S., Heathcote, L., & Poole, G. (2010). Harnessing ICT potential. International Journal of Educational Management.

[76] Vengroff, R., & Bourbeau, J. (2006, February). In-class vs. On-line and Hybrid Class Participation and Outcomes: Teaching the Introduction to Comparative Politics Class. In annual meeting of the APSA Teaching and Learning Conference.

[77] Dutt, A., & Ismail, M. A. (2019, June). Can we predict student learning performance from LMS data? A classification approach. In 3rd International Conference on Current Issues in Education (ICCIE 2018) (pp. 24-29). Atlantis Press.

[78] Lust, G., Vandewaetere, M., Ceulemans, E., Elen, J., & Clarebout, G. (2011). Tool-use in a blended undergraduate course: In Search of user profiles. Computers & Education, 57(3), 2135-2144.

[79] Hung, J. L., & Zhang, K. (2012). Examining mobile learning trends 2003–2008: A categorical meta-trend analysis using text mining techniques. Journal of Computing in Higher education, 24(1), 1-17.

[80] Damianov, D. S., Kupczynski, L., Calafiore, P., Damianova, E. P., Soydemir, G., & Gonzalez, E. (2009). Time spent online and student performance in online business courses: A multinomial logit analysis. Journal of Economics and Finance Education, 8(2), 11-22.

[81] Baugher Varanelli Weisbord, D. A. E. (2003). Student hits in an internet-supported course: How can instructors use them and what do they mean?. Decision Sciences Journal of Innovative Education, 1(2), 159-179.

[82] Biktimirov, E. N., & Klassen, K. J. (2008). Relationship between use of online support materials and student performance in an introductory finance course. journal of education for business, 83(3), 153-158.

[83] Coldwell, J., Craig, A., Paterson, T., & Mustard, J. (2008). Online students: Relationships between participation, demographics and academic performance. Electronic journal of e-learning, 6(1), 19-30.

[84] Greenland, S. J. (2011). Using log data to investigate the impact of (a) synchronous learning tools on LMS interaction. In Proceedings of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE) Conference.

[85] Nandi, D., Hamilton, M., Harland, J., & Warburton, G. (2011, January). How active are students in online discussion forums?. In Proceedings of the Thirteenth Australasian Computing Education Conference-Volume 114 (pp. 125-134).

[86] Shum, S. B., & Ferguson, R. (2012). Social learning analytics. Journal of educational technology & society, 15(3), 3-26.

[87] Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. Computers in Human Behavior, 47, 157-167.

[88] Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (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, 542-550.

[89] Dominguez, 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 EDM (pp. 589-590).

[90] Lerche, T., & Kiel, E. (2018). Predicting student achievement in learning management systems by log data analysis. Computers in Human Behavior, 89, 367-372.

[91] Reppert, S. M., & Weaver, D. R. (2002). Coordination of circadian timing in mammals. Nature, 418(6901), 935-941.

[92] Dibner, C., Schibler, U., & Albrecht, U. (2010). The mammalian circadian timing system: organization and coordination of central and peripheral clocks. Annual review of physiology, 72, 517-549.

[93] Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., & Zoran, M. J. (2005). Circadian rhythms from multiple oscillators: lessons from diverse organisms. Nature Reviews Genetics, 6(7), 544-556.

[94] Roenneberg, T., Wirz-Justice, A., & Merrow, M. (2003). Life between clocks: daily temporal patterns of human chronotypes. Journal of biological rhythms, 18(1), 80-90.

[95] Wittmann, M., Dinich, J., Merrow, M., & Roenneberg, T. (2006). Social jetlag: misalignment of biological and social time. Chronobiology international, 23(1-2), 497-509.

[96] Korczak, A. L., Martynhak, B. J., Pedrazzoli, M., Brito, A. F., & Louzada, F. M. (2008). Influence of chronotype and social zeitgebers on sleep/wake patterns. Brazilian Journal of Medical and Biological Research, 41, 914-919.

[97] Vitale, J. A., Roveda, E., Montaruli, A., Galasso, L., Weydahl, A., Caumo, A., & Carandente, F. (2015). Chronotype influences activity circadian rhythm and sleep: differences in sleep quality between weekdays and weekend. Chronobiology international, 32(3), 405-415.

[98] Carskadon, M. A., Wolfson, A. R., Acebo, C., Tzischinsky, O., & Seifer, R. (1998). Adolescent sleep patterns, circadian timing, and sleepiness at a transition to early school days. Sleep, 21(8), 871-881.

[99] Crowley, S. J., Acebo, C., & Carskadon, M. A. (2007). Sleep, circadian rhythms, and delayed phase in adolescence. Sleep medicine, 8(6), 602-612.

[100] Díaz-Morales, J. F., & Sorroche, M. G. (2008). Morningness-eveningness in adolescents. The Spanish journal of psychology, 11(1), 201-206.

[101] Baehr, E. K., Revelle, W., & Eastman, C. I. (2000). Individual differences in the phase and amplitude of the human circadian temperature rhythm: with an emphasis on morningness–eveningness. Journal of sleep research, 9(2), 117-127.

[102] Randler, C. (2007). Gender differences in morningness–eveningness assessed by self-report questionnaires: A meta-analysis. Personality and individual differences, 43(7), 1667-1675.

[103] Valladares, M., Ramírez-Tagle, R., Muñoz, M. A., & Obregón, A. M. (2018). Individual differences in chronotypes associated with academic performance among Chilean University students. Chronobiology international, 35(4), 578-583.

[104] Porcheret, K., Wald, L., Fritschi, L., Gerkema, M., Gordijn, M., Merrrow, M., ... & Foster, R. G. (2018). Chronotype and environmental light exposure in a student population. Chronobiology international, 35(10), 1365-1374.

[105] Putilov, A. A., Marcoen, N., Neu, D., Pattyn, N., & Mairesse, O. (2019). There is more to chronotypes than evening and morning types: results of a large-scale

community survey provide evidence for high prevalence of two further types. Personality and Individual Differences, 148, 77-84.

[106] Torsvall, L., & Åkerstedt, T. (1980). A diurnal type scale: construction, consistency and validation in shift work. Scandinavian journal of work, environment & health, 283-290.

[107] Smith, C. S., Reilly, C., & Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. Journal of Applied psychology, 74(5), 728.

[108] Putilov, A. A., & Putilov, D. A. (2005). Sleepless in Siberia and Alaska: Cross-validation of factor structure of the individual adaptability of the sleep-wake cycle. Ergonomia, 27(3), 207-226.

[109] Putilov, A. A., & Onischenko, M. A. (2005). The 52-item sleep-wake pattern assessment questionnaire: Evaluation of its psychometric features. Ergonomia, 27(2), 131-144.

[110] Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: circadian rhythms in human cognition. Cognitive neuropsychology, 24(7), 755-789.

[111] Roberts, R. D., & Kyllonen, P. C. (1999). Morningness–eveningness and intelligence: early to bed, early to rise will likely make you anything but wise!. Personality and Individual differences, 27(6), 1123-1133.

[112] Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. Intelligence, 14(4), 389-433.

[113] Killgore, W. D., & Killgore, D. B. (2007). Morningness-eveningness correlates with verbal ability in women but not men. Perceptual and motor skills, 104(1), 335-338.

[114] Song, J., & Stough, C. (2000). The relationship between morningness–eveningness, time-of-day, speed of information processing, and intelligence. Personality and Individual differences, 29(6), 1179-1190.

[115] Cortesi, F., Giannotti, F., Mezzalira, E., Bruni, O., & Ottaviano, S. (1997). Circadian type, sleep patterns, and daytime functioning in adolescence: Preliminary data on an Italian representative sample. Sleep Research, 26, 707.

[116] Enright, T., & Refinetti, R. (2017). Chronotype, class times, and academic achievement of university students. Chronobiology international, 34(4), 445-450.

[117] Preckel, F., Lipnevich, A. A., Schneider, S., & Roberts, R. D. (2011). Chronotype, cognitive abilities, and academic achievement: A meta-analytic investigation. Learning and Individual Differences, 21(5), 483-492.

[118] Preckel, F., & Roberts, R. (2009, September). Schulleistung und Chronotyp [School achievement and chronotype]. Paper presented at the 10th meeting of the section of Differential Psychology of the German Psychological Society, University of KoblenzLandau, Landau.

[119] Montaruli, A., Castelli, L., Galasso, L., Mulè, A., Bruno, E., Esposito, F., ... & Roveda, E. (2019). Effect of chronotype on academic achievement in a sample of Italian University students. Chronobiology international, 36(11), 1482-1495.

[120] Fancsali, S. (2015). Confounding Carelessness? Exploring Causal Relationships Between Carelessness, Affect, Behavior, and Learning in Cognitive Tutor Algebra Using Graphical Causal Models. In EDM (pp. 508-511).

[121] Fancsali, S. (2014, July). Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra. In Educational Data Mining 2014.

[122] Koedinger, K. R., McLaughlin, E. A., Jia, J. Z., & Bier, N. L. (2016, April). Is the doer effect a causal relationship? How can we tell and why it's important. In Proceedings of the sixth international conference on learning analytics & knowledge (pp. 388-397).

[123] Chen, L., Ramsey, J., Durbawski, A., (2021). Affect, Support and Personal Factors: Multimodal Causal Models of One-on-one Coaching. Journal of Educational Data Mining.

[124] Ramsey, J. D., Zhang, K., Glymour, M., Romero, R. S., Huang, B., Ebert-Uphoff, I., ... & Glymour, C. (2018). TETRAD—A toolbox for causal discovery. In 8th International Workshop on Climate Informatics.

[125] Ocumpaugh, J. (2015). Baker Rodrigo Ocumpaugh monitoring protocol (BROMP) 2.0 technical and training manual. New York, NY and Manila, Philippines: Teachers College, Columbia University and Ateneo Laboratory for the Learning Sciences, 60.

[126] Jiang, Y., Bosch, N., Baker, R. S., Paquette, L., Ocumpaugh, J., Andres, J. M. A. L., ... & Biswas, G. (2018, June). Expert Feature-Engineering vs. Deep Neural Networks: Which is Better for Sensor-Free Affect Detection?. In International Conference on Artificial Intelligence in Education (pp. 198-211). Springer, Cham.

[127] Baker, R., & Ocumpaugh, J. (2015). Interaction-based affect detection in educational software (pp. 233-245). New York: Oxford University Press.

[128] D'mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. ACM Computing Surveys (CSUR), 47(3), 43.

[129] Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., Shute, V., Ventura, M., ... & Zhao, W. (2015, March). Automatic detection of learning-centered affective states in the wild. In Proceedings of the 20th international conference on intelligent user interfaces (pp. 379-388). ACM.

[130] D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. Learning and Instruction, 22(2), 145-157.

[131] Botelho, A. F., Baker, R. S., Ocumpaugh, J., & Heffernan, N. T. (2018). Studying Affect Dynamics and Chronometry Using Sensor-Free Detectors. International Educational Data Mining Society.

[132] D'Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex learning. Cognition & Emotion, 25(7), 1299-1308.

[133] Huang, Y., Gong, J., Rucker, M., Chow, P., Fua, K., Gerber, M. S., ... & Barnes, L. E. (2017, June). Discovery of behavioral markers of social anxiety from smartphone sensor data. In Proceedings of the 1st Workshop on Digital Biomarkers (pp. 9-14).

[134] Steinhubl, S. R., Muse, E. D., & Topol, E. J. (2015). The emerging field of mobile health. Science translational medicine, 7(283), 283rv3-283rv3.

[135] d Baker, R. S., Gowda, S. M., & Corbett, A. T. (2011, June). Towards predicting future transfer of learning. In International Conference on Artificial Intelligence in Education (pp. 23-30). Springer, Berlin, Heidelberg.

[136] Hemenover, S. H. (2003). Individual differences in rate of affect change: Studies in affective chronometry. Journal of personality and social psychology, 85(1), 121.

[137] Kuppens, P., Oravecz, Z., & Tuerlinckx, F. (2010). Feelings change: Accounting for individual differences in the temporal dynamics of affect. Journal of personality and social psychology, 99(6), 1042.

[138] Mandalapu, V., & Gong, J. (2018, June). Towards Better Affect Detectors: Detecting Changes Rather Than States. In International Conference on Artificial Intelligence in Education (pp. 199-203). Springer, Cham.

[139] Shute, V. J., D'Mello, S., Baker, R., Cho, K., Bosch, N., Ocumpaugh, J., ... & Almeda, V. (2015). Modeling how incoming knowledge, persistence, affective states, and in-game progress influence student learning from an educational game. Computers & Education, 86, 224-235.

[140] Escueta, M., Quan, V., Nickow, A. J., & Oreopoulos, P. (2017). Education technology: an evidence-based review (No. w23744). National Bureau of Economic Research.

[141] Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April). Off-task behavior in the cognitive tutor classroom: when students game the system. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 383-390). ACM.

[142] Rodrigo, M. M. T., Baker, R. S., Lagud, M. C., Lim, S. A. L., Macapanpan, A. F., Pascua, S. A. M. S., ... & Viehland, N. J. (2007). Affect and usage choices in simulation problem solving environments. Frontiers in Artificial Intelligence and Applications, 158, 145.

[143] Ganganwar, V. (2012). An overview of classification algorithms for imbalanced datasets. International Journal of Emerging Technology and Advanced Engineering, 2(4), 42-47.

[144] Caruso, C., & Quarta, F. (1998). Interpolation methods comparison. Computers & Mathematics with Applications, 35(12), 109-126.

[145] Thévenaz, P., Blu, T., & Unser, M. (2000). Interpolation revisited [medical images application]. IEEE Transactions on medical imaging, 19(7), 739-758.

[146] Marchand, G. C., & Gutierrez, A. P. (2012). The role of emotion in the learning process: Comparisons between online and face-to-face learning settings. The Internet and Higher Education, 15(3), 150-160.

[147] Mierswa, I., & Wurst, M. (2006, July). Information preserving multi-objective feature selection for unsupervised learning. In Proceedings of the 8th annual conference on Genetic and evolutionary computation (pp. 1545-1552). ACM.

[148] Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. In Feature extraction (pp. 1-25). Springer, Berlin, Heidelberg.

[149] Mierswa, I., & Klinkenberg, R. (2019). RapidMiner Studio (9.2) [Data science, machine learning, predictive analytics]. Retrieved from https://rapidminer.com/

[150] Sak, H., Senior, A., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In Fifteenth annual conference of the international speech communication association.

[151] Feidakis, M. (2011). Emotional scaffolding with respect to time factors in Networking Collaborative Learning Environments. eLearn Center Research Paper Series.

[152] Kuppens, P. (2015). It's about time: A special section on affect dynamics. Emotion Review, 7(4), 297-300.

[153] Aminikhanghahi, S., & Cook, D. J. (2017, March). Using change point detection to automate daily activity segmentation. In 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops) (pp. 262-267). IEEE.

[154] Geigle, C., & Zhai, C. (2017, April). Modeling mooc student behavior with two-layer hid-den markov models. In Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale (pp. 205-208). ACM.

[155] Lent, R. W., Lopez Jr, A. M., Lopez, F. G., & Sheu, H. B. (2008). Social cognitive career theory and the prediction of interests and choice goals in the computing disciplines. Journal of Vocational Behavior, 73(1), 52-62.

[156] San Pedro, M. O., Ocumpaugh, J., Baker, R. S., & Heffernan, N. T. (2014). Predicting STEM and Non-STEM College Major Enrollment from Middle School Interaction with Mathematics Educational Software. In EDM (pp. 276-279).

[157] Hayden, K., Ouyang, Y., Scinski, L., Olszewski, B., & Bielefeldt, T. (2011). Increasing stu-dent interest and attitudes in STEM: Professional development and activities to engage and inspire learners. Contemporary Issues in Technology and Teacher Education, 11(1), 47-69.

[158] Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of edu-cational experiences with earned degrees in STEM among US students. Science Education, 95(5), 877-907.

[159] Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016, April). Combining click-stream data with NLP tools to better understand MOOC completion. In Proceedings of the sixth international conference on learning analytics & knowledge (pp. 6-14). ACM.

[160] Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In Handbook of research on student engagement (pp. 259-282). Springer, Boston, MA.

[161] Leony, D., Merino, P. J. M., Valiente, J. A. R., Pardo, A., & Kloos, C. D. (2015). Detection and Evaluation of Emotions in Massive Open Online Courses. J. UCS, 21(5), 638-655.

[162] Kim, C., & Pekrun, R. (2014). Emotions and motivation in learning and performance. In Handbook of research on educational communications and technology (pp. 65-75). Springer, New York, NY.

[163] Mierswa, I., & Klinkenberg, R. (2019). RapidMiner Studio (9.1) [Data science, machine learning, predictive analytics]. Retrieved from https://rapidminer.com/

[164] Mierswa, I. (2009). Non-convex and multi-objective optimization in data mining (Doctoral dissertation).

[165] Hall, M. A. (2000). Correlation-based feature selection of discrete and numeric class machine learning.

[166] Zhu, Y., Kwok, J. T., & Zhou, Z. H. (2018). Multi-label learning with global and local label correlation. IEEE Transactions on Knowledge and Data Engineering, 30(6), 1081-1094.

[167] Pardos, Z. A., & Heffernan, N. T. (2010, June). Modeling individualization in a bayesian networks implementation of knowledge tracing. In International Conference on User Model-ing, Adaptation, and Personalization (pp. 255-266). Springer, Berlin, Heidelberg.

[168] Idoine, C., Krensky, P., Brethenoux, E., Hare, J., Sicular, S., & Vashisth, S. (2018). Magic Quadrant for Data Science and Machine-Learning Platforms. Gartner, Inc.

[169] Cao, J., Lin, Z., Huang, G. B., & Liu, N. (2012). Voting based extreme learning machine. Information Sciences, 185(1), 66-77.

[170] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining (pp. 1135-1144).

[171] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Pro-ceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785-794). ACM.

[172] Bühlmann, P., & Hothorn, T. (2007). Boosting algorithms: Regularization, prediction and model fitting. Statistical Science, 22(4), 477-505.

[173] Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Re-search, 259(2), 689-702.

[174] Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., ... & Livak, T. (2005). The Assistment project: Blending assessment and assisting. In Proceedings of the 12th Annual Conference on Artificial Intelligence in Education (pp. 555-562).

[175] Ninoriya, S., Chawan, P. M., & Meshram, B. B. (2011). CMS, LMS and LCMS for elearning. International Journal of Computer Science Issues (IJCSI), 8(2), 644.

[176] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. Expert systems with applications, 33(1), 135-146.

[177] Plak, S., Cornelisz, I., Meeter, M., & Van Klaveren, C. (2019, March). Early Warning Systems for More Effective Student Counseling in Higher Education–Evidence from a Dutch Field Experiment. In Proceedings of the SREE Spring 2019 Conference, Washington, DC, USA (pp. 6-9).

[178] Beck, H. P., & Davidson, W. D. (2001). Establishing an early warning system: Predicting low grades in college students from survey of academic orientations scores. Research in Higher education, 42(6), 709-723.

[179] Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2016). Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. IEEE Transactions on Learning Technologies, 10(1), 17-29.

[180] Mozahem, N. A. (2020). Using learning management system activity data to predict student performance in face-to-face courses. International Journal of Mobile and Blended Learning (IJMBL), 12(3), 20-31.

[181] Riestra-González, M., del Puerto Paule-Ruíz, M., & Ortin, F. (2021). Massive LMS log data analysis for the early prediction of course-agnostic student performance. Computers & Education, 163, 104108.

[182] Chiu, Y. C., Hsu, H. J., Wu, J., & Yang, D. L. (2018). Predicting Student Performance in MOOCs Using Learning Activity Data. J. Inf. Sci. Eng., 34(5), 1223-1235.

[183] Jo, I. H., Kim, D., & Yoon, M. (2015). Constructing proxy variables to measure adult learners' time management strategies in LMS. Journal of Educational Technology & Society, 18(3), 214-225.

[184] Jo, I. H., Park, Y., Yoon, M., & Sung, H. (2016). Evaluation of online log variables that estimate learners' time management in a Korean online learning context. International Review of Research in Open and Distributed Learning, 17(1), 195-213.

[185] Gray, R. M. (2011). Entropy and information theory. Springer Science & Business Media.

[186] Kozachenko, L. F., & Leonenko, N. N. (1987). Sample estimate of the entropy of a random vector. Problemy Peredachi Informatsii, 23(2), 9-16.

[187] Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. Physical review E, 69(6), 066138.

[188] Putilov, A. A., Sveshnikov, D. S., Puchkova, A. N., Dorokhov, V. B., Bakaeva, Z. B., Yakunina, E. B., ... & Mairesse, O. (2021). Single-Item Chronotyping (SIC), a method to self-assess diurnal types by using 6 simple charts. Personality and Individual Differences, 168, 110353.

[189] Romo-Nava, F., Blom, T. J., Guerdjikova, A., Winham, S. J., Cuellar-Barboza, A. B., Nunez, N. A., ... & McElroy, S. L. (2020). Evening chronotype, disordered eating behavior, and poor dietary habits in bipolar disorder. Acta Psychiatrica Scandinavica, 142(1), 58-65.

[190] Ruping, S. (2000). mySVM-manual. http://www-ai. cs. uni-dortmund. de/SOFTWARE/MYSVM/.

[191] Hofmann, M., & Klinkenberg, R. (Eds.). (2016). RapidMiner: Data mining use cases and business analytics applications. CRC Press.

[192] Mandalapu, V., & Gong, J. (2019, January). Studying Factors Influencing the Prediction of Student STEM and Non-STEM Career Choice. In Proceedings of the 12th International Conference on Educational Data Mining.

[193] Mandalapu, V., Gong, J., & Chen, L. (2021). Do we need to go Deep? Knowledge Tracing with Big Data. arXiv preprint arXiv:2101.08349.

[194] Kennedy, G., Coffrin, C., De Barba, P., & Corrin, L. (2015, March). Predicting success: how learners' prior knowledge, skills and activities predict MOOC performance. In Proceedings of the fifth international conference on learning analytics and knowledge (pp. 136-140).

[195] Stapel, M., Zheng, Z., & Pinkwart, N. (2016). An Ensemble Method to Predict Student Performance in an Online Math Learning Environment. International Educational Data Mining Society.

[196] Kim, J. H., Seodaemun-gu, S., Park, Y., Song, J., & Jo, I. H. (2014). Predicting students' learning performance by using online behavior patterns in blended learning environments: comparison of two cases on linear and non-linear model. Korea, 120, 750.

[197] Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. Journal of Universal Computer Science, 21(1), 110.

[198] Kupczynski, L., Gibson, A. M., Ice, P., Richardson, J., & Challoo, L. (2011). The impact of frequency on achievement in online courses: A study from a south Texas university. Journal of Interactive Online Learning, 10(3).

[199] Liu, F., & Cavanaugh, C. (2011). High enrollment course success factors in virtual school: Factors influencing student academic achievement. International Journal on E-learning, 10(4), 393-418.

[200] Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. Contemporary Educational Psychology, 39(2), 145-155.

[201] Schapiro, S. R., & Livingston, J. A. (2000). Dynamic self-regulation: The driving force behind academic achievement. Innovative Higher Education, 25(1), 23-35.

[202] Toering, T., Elferink-Gemser, M. T., Jonker, L., van Heuvelen, M. J., & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). International Journal of Sport and Exercise Psychology, 10(1), 24-38.

[203] Rosenman, R., Tennekoon, V., & Hill, L. G. (2011). Measuring bias in self-reported data. International Journal of Behavioural and Healthcare Research, 2(4), 320-332.

[204] Salehian Kia, F., Hatala, M., Baker, R. S., & Teasley, S. D. (2021, April). Measuring Students' Self-Regulatory Phases in LMS with Behavior and Real-Time Self Report. In LAK21: 11th International Learning Analytics and Knowledge Conference (pp. 259-268).

[205] Ahmad Uzir, N. A., Gašević, D., Matcha, W., Jovanović, J., & Pardo, A. (2020). Analytics of time management strategies in a flipped classroom. Journal of Computer Assisted Learning, 36(1), 70-88.

[206] Landrum, B. (2020). Examining Students' Confidence to Learn Online, Self-Regulation Skills and Perceptions of Satisfaction and Usefulness of Online Classes. Online Learning, 24(3), 128-146.

[207] Ilves, K., Leinonen, J., & Hellas, A. (2018, February). Supporting self-regulated learning with visualizations in online learning environments. In Proceedings of the 49th ACM Technical Symposium on Computer Science Education (pp. 257-262).

[208] Edwards, S. H., Snyder, J., Pérez-Quiñones, M. A., Allevato, A., Kim, D., & Tretola, B. (2009, August). Comparing effective and ineffective behaviors of student programmers. In Proceedings of the fifth international workshop on Computing education research workshop (pp. 3-14).

[209] Martin, J., Edwards, S. H., & Shaffer, C. A. (2015, August). The effects of procrastination interventions on programming project success. In Proceedings of the eleventh annual International Conference on International Computing Education Research (pp. 3-11).

[210] Mandalapu, V., Chen, L., Chen, Z., Gong, J., (July, 2021). Student-Centric Model of Login Patterns: A Case Study with Learning Management Systems. 14th International Conference on Educational Data Mining.

[211] Pelleg, D., & Moore, A. W. (2000, June). X-means: Extending k-means with efficient estimation of the number of clusters. In Icml (Vol. 1, pp. 727-734).

[212] Senin, P. (2008). Dynamic time warping algorithm review. Information and Computer Science Department University of Hawaii at Manoa Honolulu, USA, 855(1-23), 40.

[213] Wang, W., Lyu, G., Shi, Y., & Liang, X. (2018, November). Time series clustering based on dynamic time warping. In 2018 IEEE 9th international conference on software engineering and service science (ICSESS) (pp. 487-490). IEEE.

[214] Hardoon, D. R., & Shawe-Taylor, J. (2011). Sparse canonical correlation analysis. Machine Learning, 83(3), 331-353.

[215] Ogarrio, J. M., Spirtes, P., & Ramsey, J. (2016, August). A hybrid causal search algorithm for latent variable models. In Conference on Probabilistic Graphical Models (pp. 368-379). PMLR.

[216] Spirtes, P., & Glymour, C. (1991). An algorithm for fast recovery of sparse causal graphs. Social science computer review, 9(1), 62-72.

[217] Spirtes, P., Glymour, C., & Scheines, R. (1993). Discovery algorithms for causally sufficient structures. In Causation, prediction, and search (pp. 103-162). Springer, New York, NY.

[218] Zavgorodniaia, A., Shrestha, R., Leinonen, J., Hellas, A., & Edwards, J. (2021, May). Morning or Evening? An Examination of Circadian Rhythms of CS1 Students. In 2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering Education and Training (ICSE-SEET) (pp. 261-272). IEEE

[219] Winne, P. H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. Contemporary Educational Psychology, 28(3), 259-276.

[220] Britton, B. K., & Tesser, A. (1991). Effects of time-management practices on college grades. Journal of educational psychology, 83(3), 405.

[221] Ghiasvand, A. M., Naderi, M., Tafreshi, M. Z., Ahmadi, F., & Hosseini, M. (2017). Relationship between time management skills and anxiety and academic motivation of nursing students in Tehran. Electronic physician, 9(1), 3678.

[222] Macan, T. H., Shahani, C., Dipboye, R. L., & Phillips, A. P. (1990). College students' time management: Correlations with academic performance and stress. Journal of educational psychology, 82(4), 760.