Educational Data Mining:

An Application of a Predictive Model of Online Student Enrollment Decisions

by

Cody Gene Singer

A Dissertation Presented in Partial Fulfillment
Of the Requirements for the Degree
Doctor of Education

Approved October 2023 by the
Gradate Supervisory Committee:

Lydia Ross, Co-Chair
Sherman Dorn, Co-Chair
David Cillay

ARIZONA STATE UNIVERSITY

December 2023

ABSTRACT

College and university enrollment has decreased nationwide every year for more than a decade as educational consumers increasingly question the value of higher education and discover alternatives to the traditional university system. Enrollment professionals seeking growth are tasked to develop and implement innovative solutions to address increasing enrollment challenges by being responsive to consumer values, interests, and needs. This multi-phase mixed methods action research study explores educational data mining and machine learning to understand and predict the enrollment decisions of admitted applicants (n=3,843) at the online campus of a public research university (phase one). Then, this innovation is distributed to understand how university enrollment professionals (n=7) interpret and are affected by the factors that influence online student enrollment decisions (phase two). Logistic regression was used to evaluate 24 independent variables to classify each applicant into a dichotomous dependent outcome: will an applicant enroll or will they not. The model identified 10 significant predictors and accurately categorized 81% of the enrollment outcomes at its peak. The population was comprised of online adult learners and the findings were carefully compared to the findings of previous studies which differed in institutional settings (on campus) and student populations (first-year students). Additionally, the study aimed to extend the work of previous literature through a second application phase within the local context. The second phase was guided by distributed leadership theory and the four-stage theory of organizational change and introduced the model to enrollment professionals within the local context through participation in a workshop coupled with a pre-/post-workshop survey. This convergent parallel mixed methods design resulted in themes that demonstrated enrollment managers had a genuine desire to understand and apply the model to assist in solving complex enrollment challenges and were interested in using the model to inform their perspectives, decision-making, and strategy development. This study concludes that educational data mining and machine learning can be used to predict the enrollment decisions of online adult students and that enrollment managers can use the data to inform the many enrollment challenges they are tasked to overcome.

ACKNOWLEDGEMENTS

The academic journey is analogous to an intricate tapestry, with numerous threads of support, encouragement, and motivation that have been woven together to create the story that tells my tale. Each chapter of this tale bears the indelible marks of those who stood by me, believed in me, and lent their unwavering support that led to accomplishments a younger me could never imagine were possible.

The first chapter of this tale is dedicated to my family, and particularly to Rebecca, my unwavering wife and best friend. Your love and sacrifice from early mornings, late nights, and countless weekends in my absence permitted me to focus on completing this journey from start to finish. I am forever grateful for the grace, resilience, and unyielding encouragement you have afforded our family.

The second chapter is a testament to my foundation and familial network. My mom, Debbie, for being the toughest woman I have ever met and instilling in me an inability to quit. My sister, Shannon, whose hard work and ultra-achiever personality has motivated me to constantly discover what I am capable of. Ken and Tina, for always showing up without being asked to help us in countless ways. Your selfless support serves as a model we aspire to achieve. And for all my family, I am so incredibly grateful to have each of you as an important part of our lives.

In the third chapter, I express my appreciation for the personal and professional mentors who have played pivotal roles in this journey. David Vande Pol, who reminds me that humility is a virtue worthy of pursuit. Your experience and wisdom have guided my trek in countless ways, and I know you won’t allow the Dr. title to make my head too big. Dr. Shari Carpenter, whose mentorship and guidance have inspired me to pursue and

persist. Your support personally and professionally has been a steadfast pillar in my life since I first stepped foot on a college campus. My committee, whose guidance and direction ushered me to the finish line. Especially Dr. Ruth Wylie, whose weekly meetings held me accountable to make progress, knowledge helped me shape my research, and support guided me. And Dr. Lydia Ross, whose relentless availability and listening ear helped me process so many of my ideas and thoughts. From the very first quantitative methods course to the final moments before submitting my dissertation, you have been a genuine advocate throughout my research. And to the many Doctors I sought for guidance and wisdom before starting this journey, especially Dr. Dixie Lund, whose bound dissertation has been on my bedside table for 4 years as a constant reminder to keep going.

The fourth chapter acknowledges the academic institutions, peers, and colleagues that structure the frame of this academic tapestry. Eastern Oregon University, my alma mater, for shaping me into who I am today and helping me start this journey. And Washington State University, for allowing me to ask difficult questions and complete important research. These organizations and the many friends and colleagues I've relied on in so many ways have been crucial to the completion of my dissertation research.

The fifth and final chapter portrays my gratitude to God and for the people placed in my life who have played a role in shaping the story that is my tale. Each thread in the tapestry of my academic journey represents a person, an institution, and a belief that has shaped my path. Your influence is eternally woven into this dissertation, and I thank you for being the pillars of support that made this possible.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION AND PURPOSE OF STUDY 1

Data Mining and Machine Learning in Higher Education...........................6

Local Context and Role of the Researcher 8

Previous Cycle of Research 10

Significance of the Study 14

Problem of Practice, Innovation, & Research Questions...........................16

2 LITERATURE REVIEW 18

Data Mining and Machine Learning in Higher Education.........................18

Adult Learners & Online Higher Education 28

Defining the Adult Learner 29

Distributed Leadership & Organizational Change: A Framework for Innovation 32

Four-Stage Theory of Organizational Change 35

Distributed Leadership Theory 36

Summary 37

CHAPTER Page

3 METHODOLOGY 39

Research Design: Mixed Methods Action Research 39

Setting 41

Phase 1: Research Questions 1 & 2 43

Research Design....................43

Population Sample 44

Instruments & Data Sources 44

Procedure 47

Data Analysis....................53

Phase 2: Research Questions 3 & 4 54

Participants....................55

Research Design....................55

Instruments & Data Sources 56

Procedure 57

Data Analysis....................59

Integration....................60

4 ANALYSIS AND RESULTS....................62

Phase One....................62

Results....................66

CHAPTER Page

Summary – Phase One 78

Phase Two 79

Pre-Workshop Survey 80

Workshop 83

Theme One: Desire to Understand 86

Theme Two: Desire for Application 88

Post-Workshop Survey 91

Summary – Phase Two 100

Summary of Findings 103

5 DISCUSSION 104

Phase One 104

Phase Two 113

Limitations 117

Recommendations for Future Research 121

Conclusion 125

REFERENCES 127

APPENDIX

A PRE/POST WORKSHOP SURVEY 134

B ONE-PAGE MODEL & PRE-SURVEY SUMMARY 140

APPENDIX Page

C WORKSHOP AGENDA..142

D MODEL FOUR – VARIABLE RESULTS 144

E QUALITATIVE CODES, CATEGORIES, AND THEMES.............................148

F IRB EXEMPTIONS…………………………………………………………...142

LIST OF TABLES

Table Page

1. Undergraduate Yield Rates by Admissions Funnel Segment …………...…..12

2. Student Variables, Data Sources, Types & Definitions……………………...45

3. Descriptive Statistics - Continuous Variables………………….………....63

4. Descriptive Statistics - Categorical Variables……………….……………....64

5. Backward Stepwise Method - Variables Removed………………………….69

6. Forward Stepwise Method - Variables Added………………………………71

7. Summary of Factors from Each Model…………………………………..…..77

8. Descriptive Statistics – Pre-Workshop Survey……………………………...82

9. Themes, Codes, & Example Statements…………………………………….84

10. Descriptive Statistics - Pre/Post-Survey Results………………….……....93

11. Dependent Samples T Test…………………………………………………..95

12. Comparison of Important Factors from Previous Studies…………….…106

LIST OF FIGURES

Figure Page

1. Definitions of Admissions Funnel Stages…..…………………………...…4

2. Applying Data Mining to the Design of Education Systems…...……........…20

3. Cross-Industry Standard Process for Data Mining (CRISP-DM)……….......26

4. Multi-Phase MMAR Design Illustration……………………..…………...41

5. Enrollment Yield by Application Month……………………..……………...74

6. Percentage Selected in Top 10 Variables…………………...…………….....98

CHAPTER 1:

INTRODUCTION AND PURPOSE OF STUDY

From 1970 through 2010, the total number of post-secondary students across the United States more than doubled (U.S. Department of Education, 2021). Since 2010, college and university enrollment has steadily declined annually nationwide. Over the last decade, institutions of higher education have shifted away from the historical model of passively enrolling students toward actively recruiting them as a response to these nationwide enrollment trends and the introduction of alternative forms of education (Roueche & Roueche, 2000). Recent innovations in education, coupled with rising tuition costs and increasing critiques about the value proposition of higher education, have led consumers to explore educational alternatives (Garrett, 2021). In the last decade, Massive Open Online Courses (MOOCs), low-cost private for-profit colleges, and free or low-cost online learning platforms like LinkedIn Learning, Coursera, EdX, and Khan Academy, have made education instantaneously available, affordable, and accessible to nearly all global consumers. In a world where you can order almost anything online and have it delivered to your doorstep within 24 hours, it is no wonder consumers are migrating away from the traditional higher education model. The historical approach to passive enrollment management with limited start terms, high tuition costs, and selective inclusivity no longer aligns with the values educational consumers have come to expect in the twenty-first century (Garrett, 2021).

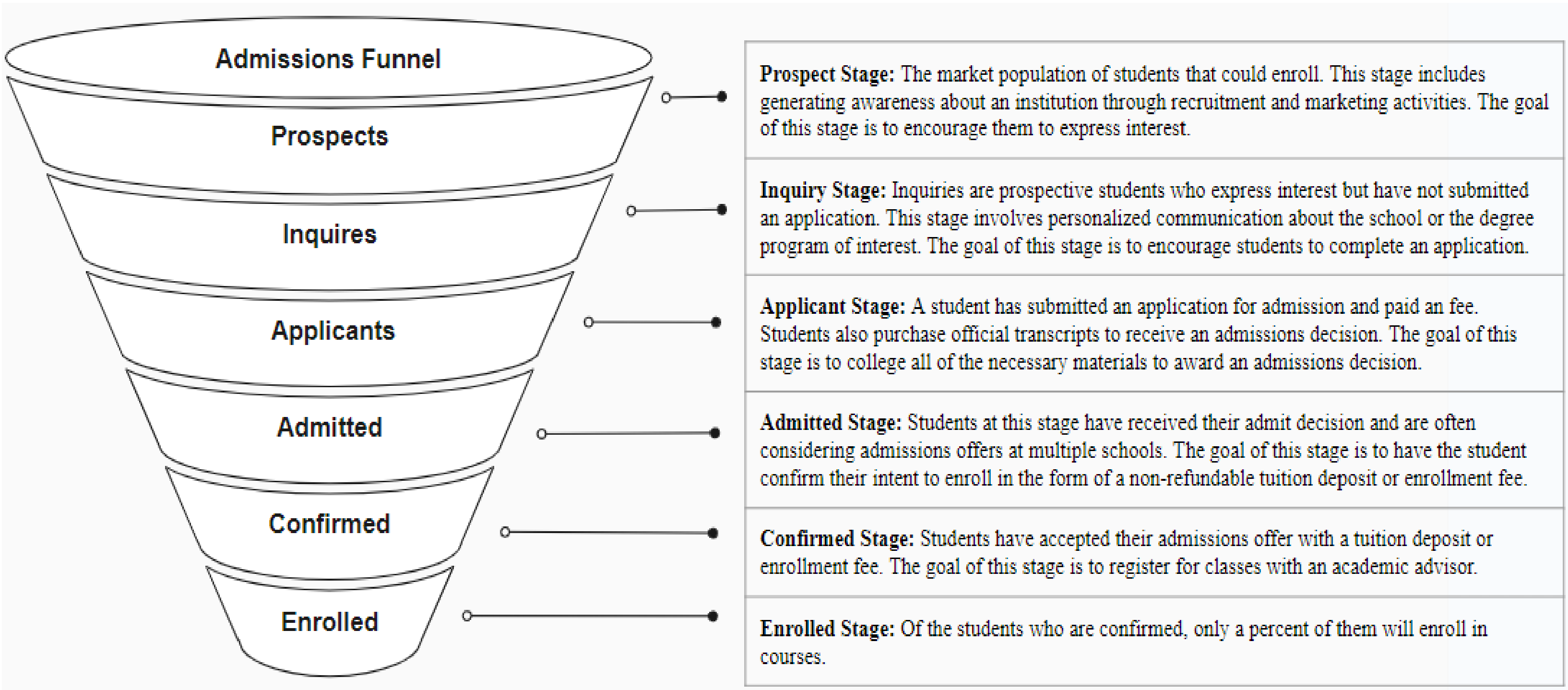
Critical scholars are increasingly questioning the value of higher education on a global scale (Blanco Ramirez & Berger, 2014; Jones, 2013; Tomlinson, 2018). The traditional educational model, as a business model, has been disrupted by innovation and has forced colleges and universities that are interested in long-term sustainability to reevaluate their place in the knowledge economy. Consumer shifts to educational alternatives have led to national declining enrollment, and consequentially, declining tuition revenue. Coupled with historic disinvestment in higher education on national and state levels (Fischer & Ellis, 2021; Knox, 2019), the sustainability of the traditional higher education model is becoming increasingly threatened. University administrators are tasked to develop and implement innovative solutions to address increasing enrollment challenges.

Institutions interested in achieving enrollment sustainability and even seeking growth must become responsive to consumer values, interests, and needs while taking a proactive stance on communicating their value proposition at the very beginning of a prospective student's journey to higher education. Perhaps more important than just generating awareness of an institution's value proposition is generating awareness among a population of prospective students who are most likely to enroll. Enrollment professionals are tasked with developing a strong understanding of their student populations and the activities that effectively encourage students to move through each stage of an admissions funnel depicted in Figure 1.

A prospective student's first interaction with a college or university typically begins with an awareness of that institution. These prospective students represent the entire market population of students that could become enrolled students. A smaller segment of potential students become inquiries who have expressed interest and identified themselves as potential students to the college or university. Even fewer eventually apply for admission, and fewer still become admitted. Of those that are admitted, only some of them confirm their intent to enroll and even fewer actually enroll in courses.

Figure 1

Definitions of Admissions Funnel Stages



The admissions funnel is an important concept because enrollment managers are tasked with understanding why some prospective students go on to the next stage and why others do not. For example, if we can understand why an inquiry decides to apply and why others do not, we can do a better job of generating high quality inquiries from a pool of prospects that best fit our particular institution. The same is true for understanding why some admitted students eventually decide to confirm and enroll. If we can better understand their enrollment decisions, we can do a better job serving those students and assisting them in the university admissions process.

While tasked with understanding the decision-making process of prospective students, admissions professionals often rely on limited data and intuition. Enrollment managers refer to yield rates to begin deciphering the enrollment behaviors of their population of students. For example, there were 12.1 million first-time degree seeking undergraduate applications to post-secondary institutions in the Fall of 2021 across the United States (U.S. Department of Education, 2021). Of those applications, only 1.6 million applicants became admitted and enrolled, 5.7 million were admitted and did not enroll, and 4.8 million will not admitted. At a national level, we have the challenge of dealing with students applying to multiple schools when calculating yield rates. At a local level that challenge is removed. For example, an online public research university in the pacific northwest (and context for this research study) received 1,721 unique student applications in the Fall of 2021. Of those total applications, 1,303 were admitted, 1128 were confirmed, and 730 enrolled in courses. This case has an applicant-to-enrolled yield rate of approximately 42%. While yield rates vary from one institution to another, understanding why 42% of applicants eventually enroll while 58% do not is a central question to many admissions professionals and is the primary focus of this study. Accurately predicting which students will or will not eventually enroll can help institutions manage their enrollment funnel and achieve enrollment growth.

The ability to accurately predict student enrollment outcomes could have substantial impacts on national and local level decision making. Admissions professionals who are proactively recruiting prospective students from a declining population are also looking to yield rates as a measure of their performance. Additionally, seeking explanations as to why students enroll or do not enroll, what patterns may or may not exist, and whether particular groups of students enroll in particular institutions can inform decision making activities for university administration. Understanding a particular population allows administrators to forecast and plan recruitment activities, target marketing and communication efforts, forecast budget and planning activities, develop scholarship and remission strategies, engage in data-informed academic course planning, among many other areas of impact. As a result of these important institutional practices and the impact of understanding and predicting matriculation rates, this study aims to use data mining techniques and machine learning to uncover answers and generate solutions to these important questions.

Data Mining and Machine Learning in Higher Education

Data mining and machine learning are techniques designed to uncover hidden patterns in large volumes of data to extract meaningful, actionable, and predictive information to inform leadership decision making (Basu et al., 2019; Chang, 2006; Luan, 2002). Data mining as defined by Schneier (2015) refers to “the science and engineering of extracting useful information from data” (p. 33). Educational Data Mining (EDM) can be defined as a discipline focused on developing methods for exploring large-scale data from educational sources with the intent to better understand and serve students (International Educational Data Mining Society, 2022). Meanwhile, machine learning can be defined as a subset of data mining and uses existing data and computer aided statistical analysis to model, identify patterns, extract useful information, and make predictions (Yates & Chamberlain, 2017).

While data mining and machine learning are well known practices in the business world and private sector, both are relatively new in their application to higher education. Nearly every targeted advertisement on social media utilizes data mining and machine learning algorithms to deliver a customized message directly to the individuals who are most likely to act or purchase a product. Higher Education's adoption of technology and use of Customer Resource Management (CRM) software and Learning Management Systems (LMS) to track and record student information generates billions of individual data points every single day. Because of the use of these technologies, we can begin to explore and answer questions that were unanswerable only 30 years ago. The field of data mining and machine learning in higher education has rapidly grown over the last two decades with the emergence of international journals (Journal of Educational Data Mining), conferences (Educational Data Mining Conference), and communities of researchers using these techniques to predict a variety of important questions. While a more detailed exploration of the literature in this field will be illustrated in chapter two, data mining and machine learning techniques have been used to predict admissions decisions (Guabassi et al., 2021; Waters and Miikkulainen, 2013), student grades outcomes (Livieris et al., 2018; Yağcı, 2022), student persistence and dropout warnings

(Nascimento et al., 2018; Yukselturk et al., 2014), evaluate teaching methods/learning outcomes (Duzhin & Gustafsson, 2018), among many other applications to answer important questions in higher education.

This study aims to introduce data mining and machine learning to answer critical questions about admissions yields in a unique setting. The remainder of this chapter will introduce the local context, the role of the researcher, and further explore the significance of this study towards the problem it aims to answer. Finally, four research questions will be introduced that outline the remainder of this dissertation and the quest to answer how machine learning and data mining can be used to predict student enrollment outcomes at colleges and universities.

Local Context and Role of the Researcher

This study is conducted as an action research study. Commonly used in education, action research refers to any systematic inquiry designed for the purpose of learning more about one's own practice to understand our practices better and improve their quality or effectiveness (McMillan, 2004; Mertler, 2020; Schmuck, 1997). Educational practitioners cite a divide between the large body of educational research literature and its practical application in an educational setting (Anderson, 2002). For this reason, action research is designed as a bridge between traditional research approaches and the development of applicable findings in local educational settings in order to improve the quality of actions and results within that setting (Schmuck, 1997). As a focused inquiry targeting a specific problem, action research allows practitioners to improve practice and improve decision-making capabilities on a local level.

As compared to the linear nature of traditional research with clear starting and ending points, action research is more cyclical in nature and is designed to build upon itself without a clear ending (Mertler, 2020). As the educational industry continually grows and our knowledge of education develops, so do the questions, challenges, and potential solutions. The process of planning, acting, developing, and reflecting on findings is a single cycle of research designed to be built upon. This allows for action researchers to utilize their findings as the basis for the next stage of action research to improve practices continually (Mills, 2011). Described later this this chapter, this dissertation is a result of a previous cycle of research.

This action research study takes place within Washington State University (WSU), a public land-grant Research I University founded in 1890. Comprised of six campuses throughout the state of Washington, WSU serves approximately 30,000 students from 50 states and 98 countries (WSU, 2022). Of the six campuses in the university system, WSU Global Campus has delivered distance and online education since 1992. Global Campus currently serves approximately 4,100 students in online bachelor's and master's degree programs around the world.

As the campus director of admissions and recruitment for WSU Global Campus, I am responsible for leading a team of admissions counselors to grow the number of new students we serve through innovative outreach and recruitment activities. My position was created in the summer of 2021 to bring leadership, strategy, and direction to grow undergraduate online enrollment. I am the first person to hold this role, and as a result, I spent previous cycles of research dedicated to learning about the position, the history of enrollment on this campus, and the opportunities available to accomplish the goal of growing online enrollment.

Previous Cycle of Research

During my first two months as the campus director of admissions and recruitment, I conducted weekly one-hour semi-structured individual interviews with three admissions counselors over a six-week period. This was my first cycle of research. The purpose of these interviews was to learn about their individual perspectives, attitudes, and understanding of online enrollment while gaining an appreciation for their historical knowledge and the local context. Prior to beginning the cycle, I learned that the organizational structure had this team of admissions counselors, a role often considered to be a junior level position, supervised by an executive level position. Part of the justification for creating and hiring my new role was to build a layer of leadership to a previously very flat organizational structure. The purpose of adding a new layer of management was to provide additional leadership, strategy, and attention to the team of admissions counselors that was previously limited due to the time constraints associated with executive level leadership.

During the first cycle of research, I learned that the team of admissions counselors held perspectives of recruitment similar to the historical description of higher education provided in the introduction of the dissertation. They engaged in very little active outreach, instead, relied on students to come to them. It is important to note, however, that this perspective towards recruitment was not created out of disinterest in active recruitment. Instead, the participants expressed that they had rarely received direction or feedback to try anything other than what they had done before. In fact, every participant communicated interest in becoming more active in recruitment and outreach activities, they simply did not have the tools or direction on how to get started. Stories of struggling to 'figure it out on their own' and not feeling the support they needed to be successful were common themes across all the interviews. Similarly, themes of optimism and the desire to try new things were almost always present across the interviews. Stories of struggle often ended in opportunistic descriptions about the future and excitement about how my new role would help them accomplish their goals. This finding is important to this local context because it demonstrated an interest to move from the historically passive model of admissions and recruitment to a more active approach.

A second finding from these interviews illuminated the absence of data, tools, and resources available for the admissions counselors to use to achieve their goals. Each of the participants' job descriptions tasked them with recruiting students; however, the participants reported that they did not have access to regular reports or queries that illustrated basic information about their prospective student population. The participants shared that the reports they had access to were often outdated, unreliable, and typically did not illustrate the information they needed to do their jobs well. In fact, the participants had no sense of how they were doing (yield rates) and where they should focus their energy, which resulted in their passive approach to recruitment. This lack of access to basic recruitment information and the participants' desire to take a more active approach to recruitment were two major research findings that informed the launch of several projects, including the one outlined in this dissertation.

Illustrated in Table 1, the WSU Department of Institution Research (2022) provides public data which explains that Global Campus admitted 74.9% of the students who applied to fall undergraduate programs between 2020-2022. Meanwhile, 88.3% of admitted students confirm their enrollment intentions with a $200 nonrefundable tuition deposit. Of those who paid the deposit, only 65.6% of them actually enrolled. That implies that 34.3% of students who paid a $200 tuition deposit and confirmed their intention to attend later decided not to enroll in courses. As a result, this study seeks to understand the student enrollment decisions of the populations of students who are admitted and enroll/do not enroll in courses.

Table 1

Undergraduate Yield Rates by Admissions Funnel Segment

Funnel Segment			**Yield Rate (Fall 2020-2022)**
Applied	to	Admitted	74.93%
Admitted	to	Confirmed	88.29%
Confirmed	to	Enrolled	65.64%
Admitted	to	Enrolled	57.95%

Focusing only on the admitted student population, 58% of them enrolled in courses and become registered students. The remaining 42% of admitted students decided not to enroll. The characteristics of both populations (admitted applicants who enrolled versus those that did not enroll) are central to the research questions proposed in this study in order to understand and potentially impact admissions enrollment strategies.

Other descriptive data important to the context of this study include student demographics. As a campus offering online undergraduate degree programs, our student population is comprised primarily of working adults, non-traditional, and transfer students. For the sake of this study, working adults are defined as individuals with part-time or greater work responsibilities while balancing part-time or greater coursework. Non-traditional students are individuals older than 25 years old who are returning to college after a prolonged absence from high school. And finally, transfer students are those who are bringing with them more than 26 semester credits (39 quarter credits) of previously completed college work. The majority of the WSU Global Campus student population belongs to one or more of these categories (working adults, non-traditional, and/or transfer students). In the three-year segment of new students evaluated in this study, applicants were, on average, 29 years old, 76% transfer students with an average of 51.6 transferable credits, 69.8% female, 77.8% Washington residents, 36.1% minorities, and 35.8% first-generation.

These descriptive statistics are helpful for the purpose of describing generalities about the student population. However, they have little utility in predicting the outcomes of our student recruitment activities in their current form. For that, I must take a deep dive into the individual characteristics of our new students. Many of these characteristics include accessible data such as their geographic locations, the timing of submitting their application, ethnicity, student type, degree type, and age among many other factors that may play a role in their likelihood of becoming a student. Data mining and machine learning permit researchers to consider hundreds of factors that may influence enrollment decisions and quantitatively measure the impact of each of those factors. Moving beyond the passive enrollment history of higher education and adopting active recruitment strategies requires admissions officers to thoroughly understand the patterns of behaviors, characteristics, needs, and values of the student populations they serve. Only then can admissions professionals focus efforts and inform decision making activities that improve admissions yield rates and target audiences of students whose needs and values most align with the value proposition our institution has to offer them. Data mining and machine learning allow for the discovery of hidden patterns and underlying relationships in data that generate models capable of predicting student yield rates, enrollment outcomes, and improving the decision-making capabilities of university leadership.

Significance of the Study

The increase in recent literature applying data mining and machine learning to problems of practice in higher education shows this method has earned recognition as a reliable and valid approach to answering challenging questions (Basu et al., 2019). Studies have used these techniques most frequently to examine data from active students. Academic and teaching data have been used to determine the effects of different teaching methods (Duzhin & Gustafsson, 2018). Student academic performance data has been used to predict student progress and pass/fail rates in several foundational studies (Livieris et al., 2018; Sekeroglu et at., 2019; Tampakas et al., 2018). Academic data has also been used to predict the likelihood of dropout or retention as an intervention early alert system in various studies (Delen, 2011; Nascimento et al., 2018; Yukselturk et al., 2014).

There are fewer examples within literature that are focused on the admissions side of enrollment management. Several private companies claim to use machine learning to help prospective students predict the likelihood of their acceptance into various colleges and universities (niche.com, go4ivy.com, etc.). These websites differ from the purpose of this study first because they do not disclose the methods used to make these predictions. Second, they are focused on the student user to predict the likelihood of acceptance, while this study is focused on using applicant data to predict the likelihood that a particular applicant will become a student. The user end of this study is focused on administrators and decision makers and the methods to answer each research question are made transparent.

Similar to the websites, other studies used machine learning algorithms to predict the likelihood an admissions committee would admit graduate student applicants based on their application data (Guabassi et al., 2021; Waters & Miikkulainen, 2013). While these studies are closer to the proposed end users of admission officers, the prediction of an admissions decision is not the purpose of this study. Instead, this dissertation will focus on predicting the likelihood that an admitted student will enroll.

The closest studies that exist were conducted to predict whether an admitted student will enroll or not using data mining and machine learning in traditional college campuses with traditional-aged freshmen applicants (Basu et al., 2019; Chang, 2006; Luan, 2002). Luan (2002) provided the foundation for admissions-based machine learning for later works completed by Chang (2006) and Basu et al. (2019). Luan (2002) conducted a large study using machine learning techniques to evaluate transfer coursework data from community colleges to universities in the Cal State system to predict which community college students will transfer. Chang (2006) applied three machine learning techniques to predict yield rates of admitted students in a large public university which was comprised entirely of first-year students between the ages 17-20 that applied to attend a physical campus. Basu et al (2019) built on Chang’s work and used seven machine learning techniques to move from yield rate predictions to individual student predictions at a small liberal arts college with traditional-aged first-year students that also applied to a physical campus.

These examples of machine learning in higher education illustrate the usefulness of this approach to answering challenging and important questions. However, each of these examples differ from the problem of practice proposed in this study, primarily in institutional settings and student populations. Existing studies focus on traditional-aged students on physical campuses while this study will expand the available literature to include online, non-traditional students on non-physical campuses. Additionally, these existing studies provide evidence that the machine learning approach is useful, reliable, and valid (Basu et al., 2019; Chang, 2006; Luan, 2002). However, these studies only provide suggestions on how the findings could be used. Another important contribution of this work is that it looks not only at the models that are created but also how the models can be used by enrollment professionals within a local context. Thus, this study will expand existing literature and bridge the gap between theories and application by disseminating the innovation to enrollment professionals throughout the organization to measure its impact on leadership decision making.

Problem of Practice, Innovation, & Research Questions

To continue to build upon the work of those researchers before me and to fill an existing gap in the available literature, this unique problem of practice can be defined as an application of data mining and machine learning to predict adult student enrollment decisions at the online campus of a public research university. Developing a predictive model using machine learning techniques and sharing the findings among key decision makers within the local context will act as the action research innovation. The distribution of the innovation among enrollment management stakeholders will be guided by Distributed Leadership Theory (Gronn, 2002; Parry & Bryman, 2006) and the Four-Stage Theory of Organizational Change (Glanz et al., 2008). The findings will measure how the models inform leadership decision making, admissions prioritization, and future research cycles through the model's cyclical improvement. This problem of practice, and the remainder of this dissertation, is designed to answer the following research questions:

> **RQ1:** Using data mining, which factors and to what extent do those factors influence the enrollment decisions of online students admitted to a public research university?
>
> **RQ2:** Using machine learning, how and to what extent do those factors predict the enrollment decisions of online students admitted to a public research university?
>
> **RQ3:** How do university enrollment professionals interpret the factors that influence online student enrollment decisions and the factors that do not?
>
> **RQ4:** How does the knowledge of factors that influence online student enrollment decisions affect enrollment professionals' perspectives, decision making, and strategy development?

CHAPTER 2

LITERATURE REVIEW

Chapter 1 provided the context to this action research study on national and local levels as well as my positionality as the researcher. It introduced example studies that this dissertation is built from in order to provide a glimpse into what is possible and illustrate the importance and originality of this particular study. In this chapter, I will present a deeper exploration of the historical literature on data mining and machine learning in higher educational settings to provide a clear understanding of how this study is informed by previous inquiries. Additionally, literature pertaining to adult learners in online undergraduate programs will be explored and linked to the absence of this population in existing data mining and machine learning literature. This gap in the literature and the unique characteristics of online adult learners influence the methodological approach and, therefore is an important section of this literature review. Distributed leadership theory and the four-stage theory of organizational change are explored and linked to the study's context and the purpose of the research to illustrate the methodological choices and theoretical perspectives. Finally, a discussion of educational data mining, adult learners in online education, and the change theories selected will introduce the innovation and the methodological approach explored in chapter three.

Data Mining and Machine Learning in Higher Education

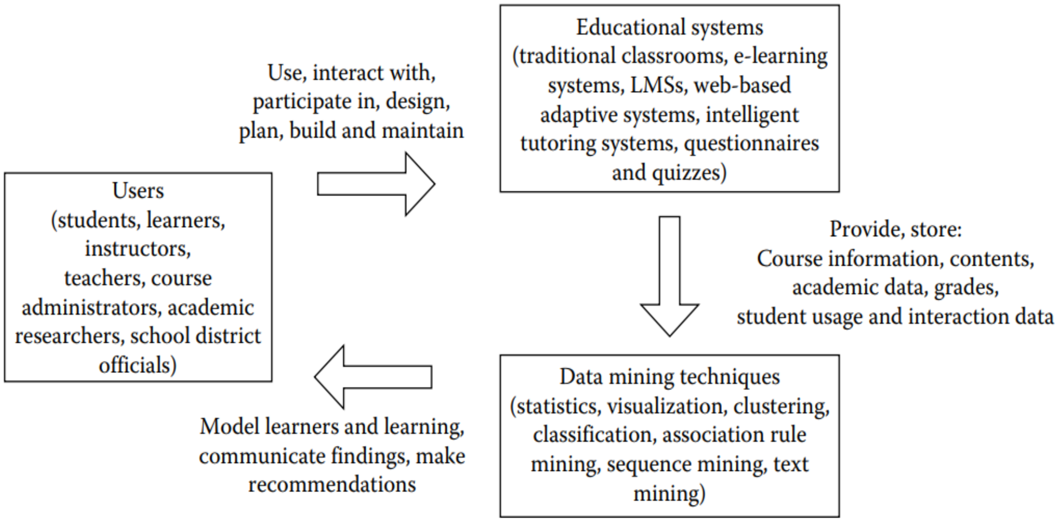
Data mining is centered on discovering patterns in large amounts of data that explain phenomena, help researchers create new knowledge, and inform future practice (Al-Twijri & Noaman, 2015; Romero & Ventura, 2007). Meanwhile, machine learning is a method to discover hidden patterns in large volumes of data to extract meaningful, actionable, and predictive information to inform leadership decision making (Basu et al., 2019; Chang, 2006; Luan, 2002). Machine learning is a predictive approach to data mining to help users make sense of possible phenomena that are otherwise inaccessible without the use of these tools.

The field of educational data mining (EDM) is fairly new and has established a growing international community of researchers since the mid-1990s (Baker & Yacef, 2009; Romero & Ventura, 2007). The *Journal of Educational Data Mining* (JEDM) was first published in 2009 and led to the first Handbook of Educational Data Mining (Romero et al., 2010) and the development of the International Educational Data Mining Society (IEDMS) in 2011.

As an emerging discipline, EDM is built from other more established data mining fields including commerce and biology (Romero et al., 2010). The discovery and use of these methods were established by researchers in higher education, but the application of data mining to education is still relatively new as compared to its use in other fields. For example, the domain of commerce is well-known for using data mining to establish buyer profiles and influence purchase decisions to drive profit. The ads on social media and other places are a result of data mining and machine learning algorithms targeting customers based on their interests and digital footprints. Meanwhile, the domain of EDM is primarily focused on improving student learning and informing administrative decision making. As an applied and formative field of research, EDM is closely tied to the methodological characteristics of action research. The cyclical nature of EDM as a field is visualized in Figure 2.

Figure 2

Applying Data Mining to the Design of Education Systems



(Romero et al., 2010, p. 3)

While much of Romero et al. (2010)'s application of data mining to the field of education is strongly informed by academic data, the Handbook of Educational Data Mining contains several case studies illustrating the use of EDM for the purposes of informing non-academic decision-making in higher education. A portion of the EDM cases that focus on administrators as stakeholders utilize machine learning techniques as predictive tools. Further, a subset of those cases that focus on predictive elements of enrollment management specifically look at admissions contexts. Since that is the focus of this study, the remainder portion of this section will explore those studies in detail as influential foundations for this particular study.

In one of the very first studies of its kind, Bruggink and Gambhir (1996) used statistical models to predict admissions decisions made by a committee and applicant enrollment decisions to a selective undergraduate liberal arts college. Focusing on characteristics important to selective colleges (academic performance, personal traits, and extracurricular activities), the researchers developed a model that correctly predicted an admissions decision 87.5% of the time while correctly predicting an applicant's decision whether or not to enroll 78% of the time.

However, what was noticeably missing from this study was the rate of acceptance/rejection, which could skew the model's predictive ability. The researchers describe the school as highly selective but do not provide admissions rates. This is a critically important omission. For example, if the school accepted 20% of their applicants and rejected 80%, then the model could achieve 80% accuracy simply by rejecting all applicants. If that were the case, correctly predicting an outcome 87.5% of the time may not result in a meaningful prediction. The researchers did not address whether they accounted for the unbalanced data. It is unclear if this omission set a precedent for future researchers who built their studies from this foundational study, but nearly all of the following studies address the issue of unbalanced data.

Waters and Miikkulainen (2013) used machine learning and a logistic regression to predict if an admissions committee at the University of Texas would admit an individual applicant (n=588) to a Ph.D. program based on the information provided in their admissions file. They created the model to rank applicants in order based on their likelihood of being admitted to the program. Then, the study ran the traditional human review process simultaneously with the machine learning processes and discovered the model made the same decisions as the human review process 87.1% of the time. The researchers addressed the data imbalance problem using precision-recall characteristics and found nearly zero false negatives, confirming the model's predictive ability was accurate. The primary outcome of this research was a demonstration that machine learning can have similar predictive abilities to human reviewers. Additionally, using machine learning permits substantial time savings by ordering files to minimize time spent reviewing rejected applicants.

Chang (2006) utilized three machine learning techniques to predict college admissions yields of traditional-aged, campus-bound, undergraduate first-year students at a large state university. Resulting from the population, the study focused on characteristics important to traditional-aged freshmen students attending a public university including high school GPA, high school rank, high school size, and SAT/ACT scores. This research used logistic regression, neural networks, and classification/regression trees (C&RT) and found that they were able to accurately predict the enrollment decision 74%, 75%, and 64% of the time, respectively. The university admitted more than 90% of applicants, so the study was limited to just looking at admitted applicants, eliminating the unbalanced data issue. Chang (2006) further supported EDM and machine learning as appropriate methods for predicting student enrollment decisions. As a result, the author recommended administrators to target outreach messages to applicants in the middle quartiles who were neither very likely nor very unlikely to enroll. While the implementation of such a recommendation was outside the scope of this study, the recommendation assumed that ranking admitted students in quartiles would allow administrators to allocate and focus resources on students they have the highest likelihood of influencing to improve student yield rates. Additionally, Chang (2006) recommended using data to inform recruiting decisions and the allocation of recruiting budgets toward students who would most likely fit their model as ideal applicants.

An important observation about Chang’s (2006) study is the variables he decided to use as predictive variables. It is noteworthy that he omitted financial information as a predictive variable but used recruitment/outreach communication type and frequency. The author noted the omission of financial information due to its complexity seemed like a potentially missed opportunity and advocated for its use in future research. However, the use of recruitment communication was the first of its kind and proved to be useful in making predictions.

Chang (2006) also utilized demographic information including gender, ethnicity, and age. Demographic and financial information within a machine learning model poses ethical questions worth considering. For example, the data presented in the article shows that 40.4% of all admitted applicants enroll in courses. When you break enrollment yield down by race, only 27% of admitted Asian students enroll and 73% choose not to enroll. Chang (2006) suggests using this model to build a target demographic for the allocation of recruitment resources and strategies. Using demographic and financial information in a predictive model may simply reinforce the population of students a school already has and introduce potential bias into the allocation of admissions resources. Using the ethnicity example as applied to Chang’s (2006) study, a machine learning model would potentially discount students who identify their race as Asian and categorize them as potentially less likely to enroll. Therefore, creating or reinforcing the prioritization decision to disinvest in resources designed to support and recruit prospective Asian students. While demographic information has been identified by every one of these studies as useful to make statistically significant predictions, it can also introduce potential bias and limit the ethical utility of predictive models. Addressing demographic differences and how they do or do not influence a model's utility is an important step not identified in Chang's (2006) study.

Further, demographic data in machine learning can also provide value. Using the same example from the Chang (2006) study, the model could identify that Asian students enroll at a lower rate, thus informing practitioners of potential bias or a lack of support mechanisms within the admissions process for this population. Therefore, practitioners could use this information to make concentrated efforts and improve internal processes to support populations identified with low yields.

Most recently, Basu et al (2019) conducted a study designed to predict admitted applicants' admissions decisions at a small liberal arts college using seven supervised machine learning techniques. Applying campus-bound student data, this study used the same techniques as Chang (2006) and added Naïve Bays, Support Vector Machine, K-Nearest Neighbors, Random Forests, and Gradient Boosting. Additionally, this study built from Chang's (2006) foundation and used features selection to determine which independent variables provide the most predictive power and eliminate those that do not contribute to the model's performance. Features were narrowed from 35 independent variables to 15 which were selected to predict the binary outcome: will a student enroll or not enroll.

Similar to Chang (2006), Basu, et al. (2019)'s student population primarily consisted of traditional-aged students attending a physical campus. As a result, influential features included high school size and rank, high school GPA, reading academic rating, and extracurricular interests. In addition to logistic regression, Basu et al. (2019) tested seven machine learning techniques including naïve bayes, decisions trees, support vector machine, k-nearest neighbors, random forests, and gradient boasting. They found logistic regression to be the most accurate predictor of a student's enrollment decisions at 79.6% accuracy. While these studies were similar in population, Basu et al (2019) advanced the work of Chang (2006) through a more comprehensive, robust, and transparent methodological approach.

Both Chang (2006) and Basu, et al. (2019) utilized a similar framework for the development of their respective models. This framework is the standard six-step data mining procedure called the Cross-Industry Standard Process for Data Mining, or CRISP-DM, developed by Chapman, et al. (2000). Illustrated in Figure 3, the CRISP-DM framework is adopted as a methodological framework in this study.

Figure 3

Cross-Industry Standard Process for Data Mining (CRISP-DM)

1 Business Understanding	**2 Data Understanding**	**3 Data Preparation**
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria* **Assess Situation** *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits* **Determine Data Mining Goals** *Data Mining Goals* *Data Mining Success Criteria* **Produce Project Plan** *Project Plan* *Initial Assessment of Tools and Techniques*	**Collect Initial Data** *Initial Data Collection Report* **Describe Data** *Data Description Report* **Explore Data** *Data Exploration Report* **Verify Data Quality** *Data Quality Report*	**Select Data** *Rational for Inclusion/ Exclusion* **Clean Data** *Data Cleaning Report* **Construct Data** *Derived Attributes Generated Records* **Integrate Data** *Merged Data* **Format Data** *Reformatted Data* *Dataset* *Dataset Description*
4 Modeling	**5 Evaluation**	**6 Deployment**
Select Modeling Techniques *Modeling Technique* *Modeling Assumptions* **Generate Test Design** *Test Design* **Build Model** *Parameter Settings* *Models* *Model Descriptions* **Assess Model** *Model Assessment* *Revised Parameter Settings*	**Evaluate Results** *Assessment of Data Mining Results w.r.t. Business Success Criteria* *Approved Models* **Review Process** *Review of Process* **Determine Next Steps** *List of Possible Actions Decisions*	**Plan Development** *Development Plan* **Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan* **Produce Final Report** *Final Report* *Final Presentation* **Review Project** *Experience Documentation*

(Reproduced from Chapman et al., 2000)

While the previous studies add to EDM as a field and illustrate the utility of data mining and machine learning in an enrollment or admissions context, the populations, scopes, and contexts in which they were completed differ from the one proposed in this study. Bruggink and Gambhir's (1996) study included both admissions decisions and enrollment decisions for undergraduate applicants at a highly selective private college. The scope of this study focuses on undergraduate applicants at a highly inclusive public university. Waters and Miikkulainen (2013) predicted the likelihood of an admission decision for Ph.D. applicants. This study focuses on undergraduate applicants and their enrollment decision. Chang (2006) and Basu et al (2019) studied traditional-aged first-year students who applied to attend physical campuses to predict enrollment decisions. This study focuses primarily on nontraditional adult learners applying to an online public research university.

Very few studies focus on enrollment commitment decisions at accessible public universities, and none focus on adult learner populations or online undergraduate programs. The exploration of each study included some of the key predictive features related to their respective populations. Online adult learners have unique needs and experiences than the features used by previous studies. Predicting an applicant's enrollment decision depends on selecting features that best represent that unique group of applicants. Therefore, this study aims to use similar methodological approaches and the CRISP-DM framework to make enrollment predictions for a new population. As a result, the characteristics of adult learners and online undergraduate programs are the focus of the next thread of this literature review and inform many of the important methodological choices presented in chapter three.

Adult Learners & Online Higher Education

There are 40.4 million Americans who have some college credit but have never earned a degree and are no longer actively enrolled (Causey et al., 2023). That amounts to more than 12% of the population in the US that has left higher education, likely took on debt, and does not have a completed degree to show. Seventy-seven percent of those potential completers are adults between the ages of 24 and 39. The State of Washington makes up 1.8% of total enrolled students across the country and 3.2% of all individuals with some credit and no degree (n = 1.16 million). These are people in our communities who have families, hold jobs, and have responsibilities that would prevent them from returning to a traditional college campus to finish what they started and experience the promise of higher education: career mobility, socio-economic mobility, and complex critical thinking skills.

Online education provides access to a population of students who could not access these benefits otherwise. Of the more than 40 million students with some college credit and no degree, 4% of them were previously enrolled in a primarily online institution, while 12% of those who reenrolled chose a primarily online institution to return and complete their degree (Shapiro et al., 2019). While this three-fold increase in adult learners reenrolling in online institutions is substantial, it is likely even higher. The Some College, No Degree report (2023) defines 'online' as institutions that serve more than 90% of their degree programs completely online. The authors acknowledge that this classification does not consider institutions with distance and online degree programs that also serve students at physical campuses. Therefore, the authors note that the share of students returning to online degree programs is likely underestimated and the share of adult students reenrolling in online public universities is likely much higher.

High-quality undergraduate education made available online and flexible to complex student schedules removes many of the barriers for adult students to complete their education and earn their degree or credential. Online undergraduate programs make higher education accessible to adult students who otherwise would be unable to complete a degree. Since the needs of adult learners differ from the needs of traditional-aged students, this section of the literature review seeks to illustrate those differences and how they inform this study.

Defining the Adult Learner

Adult learners, historically referred to as non-traditional students, can be defined as belonging to one or more of the following groups:

1. entry or return to college at 25 years of age or older,
2. having dependents,
3. being a single parent,
4. being employed full time while enrolled,
5. being financially independent, or
6. entry is delayed by at least one year following high school (Choy, 2002; NCES, n.d.; Ross-Gordon, 2011).

In 2019, there were over 4.3 million students enrolled in postsecondary undergraduate degree seeking programs that were age 25 or older (USDE, 2021). More than 1 in 4 students enrolled nationally met the criteria of being an adult learner by age alone. While age is a common factor for membership in this group, so are the many other characteristics that are more difficult to measure. Adult learners face unique challenges as compared to their traditional student counterparts that typically include balancing multiple roles and responsibilities while attending school (Choy, 2002; Ross-Gordon, 2011). Characteristics that distinguish adult learners from others show a balancing act that often includes responsibilities for employment, to a spouse or partner, as a parent or caregiver, and as a community member. Unlike students who enroll in college immediately after high school that have the capacity to focus most of their attention on academics, adult learners are often financially independent from their parents and are primarily responsible for covering the costs of major living expenses which require the simultaneous responsibility of maintaining full time work while enrolled in courses. While the literature often paints the characteristics of adult learners as assets to the educational community due to their life experience and ability to connect theory to practice (Berker & Horn, 2003), the balancing of multiple roles often presents challenges to adult learners because time is a finite resource that must be divided among their many responsibilities.

Adult learners also have different motivations for starting or returning to complete their degree than traditional students. Traditional students often experience a period of discovery, personal growth, and maturity that a well-rounded education provides them during a formative time in which they are away from their parents and on their own for the first time. This ‘college-experience’ is sought after and considered a coming of age milestone that leads to a rapid period of student development between ages 18 and 22. While extrinsic motivators like career preparation, promotion eligibility, and earnings potential are common among both populations of students (Aslanian, 2001; Osgood-

Treston, 2001), online adult students report intrinsic motivations more frequently than their traditional-aged counterparts (Dumais et al., 2013). Intrinsic motivations include personal fulfillment, role modeling for children, and family encouragement (Dumais et al., 2013; Kimmel & McNeese, 2006).

The unique experiences, challenges, and opportunities that come with being an adult learner can be summarized through the theoretical lens of andragogy. Andragogy is an adult learning theory that attempts to describe the differences between the way adults and children learn (Corley, 2022). Developed by Malcolm Knowles (1980), andragogy posits that adult learners have distinguishing learning characteristics that influence their learning motivations and abilities. Knowles (1980) posits that adult learners have five characteristics that distinguish them from other learners. This includes the ability to direct their own learning and move away from the teacher-learner dependency experienced by traditional students. Adult learners utilize life experiences to aid learning and the application of learning materials unlike traditional learners who have limited life experience to be able to draw upon and build real-world connections. They have independently made the decision to continue learning and therefore have clear motivations and purpose for their learning whereas traditional students may continue their education because that is simply a natural next step after high school. Adult learners also have the desire to apply learning immediately to their context and is problem centered whereas traditional students may not have an immediate professional context to apply new knowledge towards. Finally, adult learners are primarily motivated by internal factors rather than external factors often experienced by traditional-aged learners (Knowles, 1980).

Andragogy as a theoretical lens is important to understanding the adult learner both in learning settings as well as the lens that drives their motivations and decisions. Understanding how adult learners experience education differs from those of traditional learners helps recruitment and enrollment management professionals cater materials and outreach efforts to address those differences and speak effectively to different audiences. Our approach to recruiting online adult learners, and therefore the factors we use to predict their enrollment decisions, differ from those used to understand traditional-aged students.

The balancing of multiple and conflicting roles coupled with the intrinsic motivational identities and andragogical characteristics are the unique elements that primarily distinguish the online adult learner from a traditional-aged campus-bound student. Previous studies aimed at predicting enrollment decisions for traditional-aged students applying to physical campuses all use similar predictive features that are appropriate for those populations (Basu et al., 2019; Chang, 2006; Luan, 2002). However, the defining characteristics of online adult learners and the factors that motivate them to enroll are often different. Therefore, these characteristics are utilized to inform the features selected in the methods chapter of this dissertation.

Distributed Leadership & Organizational Change: A Framework for Innovation

The previous sections address the foundational studies, theoretical lenses, and frameworks that inform the methods selected to answer the first two research questions. The first aimed to understand which factors and to what extent those factors influence the enrollment decisions of admitted online undergraduate students. The second was designed to understand how and to what extent those factors predict the enrollment decisions of those students. The historical development of previous studies and the literary gap in institutional setting and student population inform the importance and context of this study. The exploration of characteristics and theories of online adult learners informs the need for measuring different factors than previous iterations which has the potential to fill an existing gap within the body of literature.

Finally, this section explores the theoretical lens, frameworks, and methodological choices to address the final two research questions: How do university enrollment professionals interpret the factors that influence online student enrollment decisions and the factors that do not? And how does the knowledge of factors that influence online student enrollment decisions affect enrollment professionals' perspectives, decision making, and strategy development?

Researchers have long theorized about how change occurs within social and organizational settings. A simple library search for theories of organizational change produces thousands of results and hundreds of theories across centuries of researchers and philosophers. Within the last century, change theories have developed to become gradually less authoritative and hierarchical to become more collaborative and democratic. Some of the more well-known change theories including Lewin's Change Theory (1947) and Rogers's Diffusion of Innovation (1962) all posit that successful change can be executed by a highly skilled leader who masterfully addresses prescribed steps for organizational change. These are the most frequently cited change theories in organizational literature and provide the foundation for hundreds of theory adaptations to modern organizations and still-developing theories of change. The Distributed Leadership Framework (2000) and the Four-State Theory of Organizational Change (2008) are examples of newer theories developed from their predecessors but have focused on more collaborative approaches to facilitating change within modern complex organizational settings.

In conducting a thorough literature review and critically reading the literature on the theories mentioned above, I continued to discover that the newer a theory was, the more collaborative its paradigm. Each of the above theories are ordered chronologically based on their development and appearance in literature. Additionally, the emergence of newer theories cite the former theories as their own theoretical lenses. Starting with Lewin's (1947) change theory all the way through modern theories on deliberative inquiry, the theories become more and more collaborative over time. The foundational and highly cited theories of the mid-1900s take the perspective that individual leaders lead change and provide frameworks for helping organizations and their people successfully adopt change. The latter theories show an emerging trend in change leadership that emphasizes leadership as a collaborative and collective experience rather than an individual one (Vartto, 2019). The latter theories tend to describe change leadership, where decision-making is a collective responsibility among several individuals rather than a single manager or leader. Though relatively newer, the collaborative approach to change theories best aligns with the research questions that aim to understand how new information affects the decision-making process for a collective group of enrollment managers. As a result, this study uses the four-stage theory of organizational change and distributed leadership theory as the theoretical frameworks that guide the inquiry into research questions three and four.

Four-Stage Theory of Organizational Change

The four-stage theory of organizational change is presented by Glanz et al. (2008) and is a modern combination of Lewin's change theory and Roger's diffusion of innovation theory. Unlike those theories that are designed to fully implement an established change, stage theory of organizational change can be used as a framework to understand how the change or innovation evolves before a full implementation/institutionalization of a change occurs. The stages of this theory include:

1. awareness of a problem and possible solutions,
2. decision to adopt the innovation,
3. implementation that includes redefining the innovation and modifying organizational structures to accommodate it,
4. institutionalization of making the innovation part of the organization's ongoing activities (Glanz et al., 2008).

While the application of this theory is described in more detail in the methods chapter of this dissertation, the four-stage theory of organizational change is the framework that drives the development of a predictive model as an innovation that addresses a shared problem. A workshop was used as an innovation to collaboratively evaluate the utility of the model and its potential for adoption in the ongoing activities of the organization. In stage one, the problem is identified to understand why some students decide to enroll and others don't. The possible solutions are the results of RQ1 and 2. In stage two, a collaborative group workshop was conducted as an innovation that asked for enrollment professionals' feedback on the findings while engaging in dialog on how they believed the findings could or could not be useful in the work they are responsible for.

This intervention informs stage two to understand further if a consensus to adopt the innovation exists. Stage three asks the participants how adopting the innovation would change or require modifications of our current work practices, procedures, and structures. Finally, stage four discusses how to utilize and institutionalize the innovation as part of ongoing work activities. This theoretical framework will be used as a guide to facilitate the collaborative workshop.

Distributed Leadership Theory

Distributed Leadership Theory (DLT) is a conceptual perspective that posits the work of leadership takes place across many people in complex organizational settings, particularly those in educational settings (Bolden, 2011). Instead of focusing on the characteristics of individual leaders and their individual roles, DLT focuses on a collective of leaders and how their shared interests and collective nature are distributed across an organization. Specifically, DLT sees leadership decision making as a social process with cross-organizational impacts and should occur within the intersection of collective leaders, followers, and the situational context (Bennett et al., 2003).

While similar to Wenger's (1998) communities of practice theory, which suggests groups of people with shared interests can collaboratively learn and improve educational practice, DLT was popularized near the turn of the century by Grown (2000) and has gained considerable popularity in the last two decades. In reflecting on Grown's (2000) work, Bennett et al. (2003) contrasted the popular leadership approach of the past century with this definition of DLT:

> "Distributed leadership is not something "done" by an individual "to" others, or a set of individual actions through which people contribute to a group or organization… it is a group activity that works through and

within relationships, rather than individual action" (Bennett et al., 2003, p. 3).

Vartto (2019) further advances DLT as a methodological framework that uses special deliberative sites where participants (as a collaborative leadership) would discuss, review, and adopt a change together. Vartto (2019) suggests that within a change context, deliberative sites can allow teams to gather information about a potential organizational change, collaboratively deliberate about the change, and develop solutions that accomplish shared goals.

DLT fits the local context due to the collaborative nature of the team of enrollment managers within the problem of practice. Additionally, this theory illustrates a theoretical perspective about the collaborative approach to the study's innovation through participation in a workshop (as a deliberative site) designed to facilitate group discussion, decisions, and possible implementation across multiple stakeholders within a complex organizational setting. As a result, the combination of distributed leadership theory and four-stage theory of organizational change provides the theoretical frameworks that guide research questions three and four.

Summary

This action research study is designed to address a meaningful gap in the existing literature and uses educational data mining and machine learning to predict online adult learner enrollment decisions at a public university. Informed by foundational studies and frameworks for effectively organizing this type of study, I aimed to predict the enrollment decisions of an entirely unique population within a unique context using the CRISP-DM framework and andragogy as a lens that differentiates this population from previous studies. Additionally, distributed leadership theory and four-stage theory of organizational change were selected as theoretical lenses that fit the local context of this action research study as well as the positionality of the researcher. The exploration of foundational studies this dissertation is built from as well as the theoretical lenses in which the study is viewed illustrates its importance and originality.

This chapter presented a deeper exploration of the historical literature on data mining and machine learning in higher educational settings and provided a clear understanding of how this study is informed by previous inquiries. Additionally, literature pertaining to adult learners in online undergraduate programs was explored and linked to the absence of this population in existing data mining and machine learning literature. This gap in the literature and the unique characteristics of online adult learners became a distinguishing feature that influenced the methodological approaches and research interventions described in the following chapter.

CHAPTER 3

METHODOLOGY

This study used a mixed methods approach to collect, analyze, and extract meaning from both quantitative and qualitative data. The study was designed to understand the characteristics that inform online student enrollment decisions as well as how that information influences enrollment professionals and their approaches to admissions and recruitment activities. This chapter briefly outlines mixed methods research and reintroduces the setting and context where the study occurred. Then, a detailed description of the multi-phase research approach includes information about the study participants, the research design, instruments and data sources, and data analysis procedures. This chapter concludes with a description of how the quantitative and qualitative data were integrated to answer the research questions.

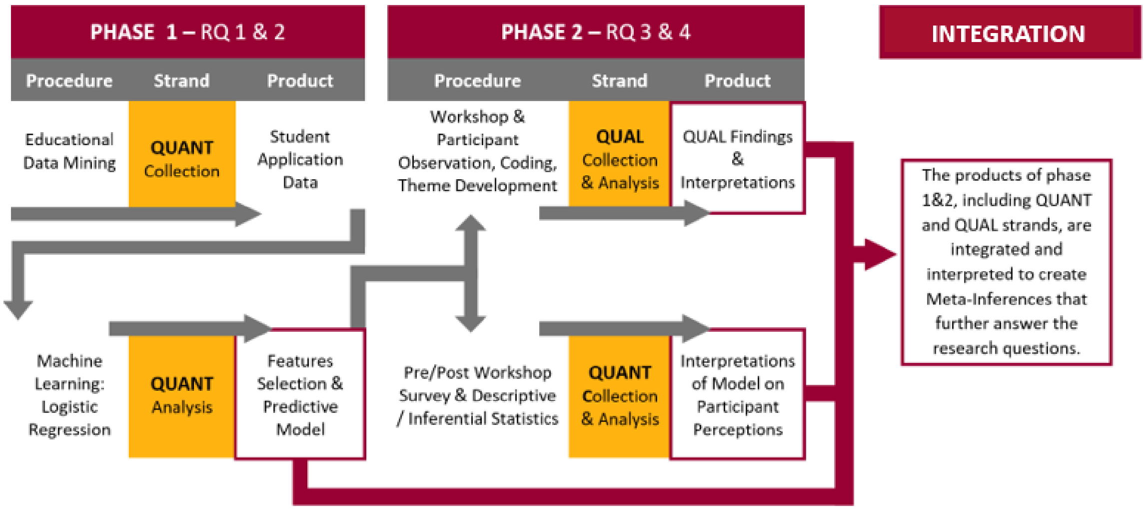
Research Design: Mixed Methods Action Research

This study utilized a multi-phase Mixed Methods Action Research (MMAR) design, which enabled the collection of quantitative and qualitative data in both separate and simultaneous phases. (Creswell & Guetterman, 2019; Ivankova, 2015). Action research cycles are often conducive to mixed method designs because the benefits of utilizing both quantitative and qualitative data can provide a better understanding of the topic than either type of data could provide by itself (Creswell, 2005; Mertler, 2020; Ivankova, 2015). Mixed methods research can be illustrated by utilizing research strands, or data collection and analysis sequencing. Methodological characteristics include the emphasis given to quantitative or qualitative methods and how the data are integrated (Creswell & Plano Clark, 2011; Ivankova, 2015; Teddlie & Tashakkori, 2009).

This multi-phase MMAR design included two data collection and analysis phases each with a unique methodological approach and a third integration stage. Structured similarly to an explanatory sequential MMAR design, phase one utilized methodological characteristics associated with quantitative studies. Specifically, data was collected using educational data mining techniques from student application records spanning three years. Then, the quantitative data was analyzed using a machine learning technique, logistic regression. The explanatory sequential design begins with a quantitative data collection and analysis followed by a qualitative strand of research (Creswell & Plano Clark, 2011). In this multi-phase MMAR design, however, phase two utilized the quantitative findings from stage one to implement a separate strand of research designed as a convergent parallel mixed methods design. Stage two concurrently implemented quantitative and qualitative strands while analyzing them separately and only combining the findings during the interpretation process. Specifically, qualitative data were collected from enrollment professionals' (participants) interactions in a workshop and open-ended survey questions while quantitative data were collected utilizing a pre- and post-workshop survey of the same participants. Qualitative data were analyzed by inductive coding, categorizing, and extracting themes from the workshop/survey responses while quantitative survey data were analyzed using descriptive and inferential statistics. In a final summary, the quantitative findings from phase one were integrated with the quantitative and qualitative findings from phase two to provide a discussion and interpretation of the study's overall findings. Figure 4 illustrates the methodological approach to this multi-phase mixed methods action research design.

Figure 4

Multi-Phase MMAR Design Illustration



Setting

As mentioned in Chapter 1, this study took place at Washington State University (WSU), a public land-grant Research I University founded in 1890. Comprised of six campuses throughout the state of Washington, WSU serves nearly 30,000 students from 50 states and 98 countries (WSU, 2022). Of the six campuses in the university system, WSU Global Campus has delivered distance and online education since 1992. Global Campus currently serves approximately 4,100 students in online bachelor's and master's degree programs around the world.

As campus director of admissions and recruitment for WSU Global Campus, I am tasked to grow new student enrollment and lead innovative outreach and recruitment activities with a team of admissions counselors. This position was created in the summer of 2021, and I was the first person to hold this role. As a result, I spent a previous cycle of research dedicated to learning about the position, the history of enrollment on this campus, and the opportunities available to grow online enrollment.

As described in chapter one, 57.1% of students admitted to WSU Global Campus went on to enroll in courses while the remaining 42.9% chose not to enroll. The dependent variable for this study was the enrollment outcome of admitted students. The student population is comprised of primarily working adults, non-traditional, and transfer students enrolled in online undergraduate degree programs. These students balance part-time or more work responsibility with coursework, are mostly over the age of 25, are returning to college after a prolonged absence, and are transferring more than 26 semester credits of previously completed coursework. The vast majority of students in the three-year population segment meet one or more of these and were, on average, 29 years old, 76% transfer students, 69.8% female, 77.8% Washington residents, 36.1% minorities, and 35.8% first-generation.

Descriptive statistics help us describe generalities about the population and the context but do not help predict the outcomes of student enrollment decisions in their current form. Various variables must be transformed using machine learning to further understand and predict the enrollment decisions of this unique population. Hidden patterns in large volumes of data can be illuminated when applying the principles of data mining and machine learning which may lead to predictive models of student enrollment outcomes and eventually improve the decision-making capabilities of university leadership.

The research was designed in two primary phases described in detail in the following sections. Phase one was a quantitative study designed to answer the first two research questions using principles of educational data mining and logistic regression as a machine learning tool to build a model to predict student enrollment decisions. Phase two introduced this innovation to enrollment professionals within the local context through participation in a workshop. Using both quantitative and qualitative instruments, phase two utilized a pre/post workshop survey in addition to participant observation to collect and analyze data produced by the workshop. Phase two was designed to answer research questions 3 and 4.

Phase 1: Research Questions 1 & 2

Phase one was designed to answer the following research questions:

RQ1: Using data mining, which factors and to what extent do those factors influence the enrollment decisions of online students admitted to a public research university?

RQ2: Using machine learning, how and to what extent do those factors predict the enrollment decisions of online students admitted to a public research university?

Research Design

While both research phases were developed using the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework introduced in Figure 3, phase one covers the first five steps of this six-step framework. This included illustrating a clear business purpose, the collection, description, and verification of data, the preparation and cleaning of data, developing the model, and evaluating the results of the model. Step five is further implemented in phase two and step six, deploying the model, is completely implemented in phase two of the research.

Population Sample

This study used data from all undergraduate students who applied and were admitted to Washington State University Global Campus during the Fall semester from 2020 to 2022 (n = 3843). Over this five-year period, 2196 admitted applicants chose to enroll while 1647 did not enroll, resulting in a three-year admit/enroll yield rate of 57.14%. All participants were coded with a 1 (enrolled) or 0 (did not enroll) as the dependent variable. A series of independent variables pulled as data from data sources are explained in the Instruments & Data Sources section below.

Instruments & Data Sources

Before data collection, a list of desired variables was compiled based on factors evaluated in previous research. During the data collection process, it became clear that some of the factors I was interested in using were not easily accessible. For example, sensitive information like financial aid data was not readily available without special permissions I was unable to receive access to. Table 2 illustrates the data I was able to access during the study period. The data was collected and verified using multiple sources including census data from the WSU department of Institutional Research (IR), a Customer Resource Management (CRM) program used for managing the student experience data, and WSU's Student Information System (SIS). The variables provided by each report are defined in Table 2 and the data is further organized into data types that include binary (BI), categorical (CA), and numerical (NU). The IR reports were built over the course of five weekly meetings, for 2 hours per week, with six data professionals from IR. Some of this data was used to create additional data sets via features engineering. For example, the variable application date submitted was subtracted from a known application deadline date to create a new variable, # of days application was submitted before the deadline. CRM Data was collected in collaboration with WSU Global Campus's Director of Enrollment Management and the CRM Program and Data Visualization Manager over the course of several months. This data came from application and student orientation reports. Finally, the SIS data was collected in collaboration with all the above contributors.

Table 2

Student Variables, Data Sources, Types & Definitions

Factor	Data Source	Data Type	Definition
Enrollment Indicator	IR	BI	Indicates whether an applicant enrolled (1) or did not enroll (0) – Dependent Variable
Age	IR	NU	Applicant's age at the time of application
Ethnic Origin Description	IR	CA	White, American Indian/Alaskan Native, Asian, Black/African American, Hispanic/Latino, Native Hawaiian/Pacific Islander, Two or more races, Not reported.
First-Generation Status	IR	BI	Indicates whether an applicant's parents completed a college level degree (1) or have not (0). Applicants whose parents have not completed a college degree (0) are considered 'first-generation'
Gender	IR	BI	An applicant's reported sex as female (1) or male (0)
Military Affiliation	SIS	BI	Indicates whether an applicant is/was a member of the military (1) or not (0)
Application Major (Degree Name)	SIS	CA	Degree program (Major) selected at the time of application

Cumulative Converted GPA	IR	NU	The cumulative GPA of previous High School or College work used for an admissions decision
Degree Type	IR	CA	First Undergraduate Degree, Post-Baccalaureate, or Non-Degree Seeking
Student Type Code	IR	CA	Former Student Returning (FSR), First-Year Student (FRS), Transfer Student (TRN), Non-Degree Student (NDG)
Total Transfer Credits	IR	NU	The total number of college-level transferable credits transferred to WSU
Last School Attended	IR	CA	The name of the last school attended (high school or college name)
City	IR	CA	The name of the city in the applicant’s physical address
County	IR	CA	The name of the county in the applicant’s physical address
State	IR	CA	The name of the state in the applicant’s physical address
Physical Address Zip Code	IR	CA	Applicant's physical address zip code
Washington (WA) Residency	IR	BI	Indicates whether an applicant is a WA resident (1) or is not a WA resident (0) at the time of application
Distance from Flagship Campus (Miles)	IR	NU	The total number of miles from the applicant's physical address zip code to 99164 (Zip code of Pullman, WA)
Application Semester	IR	CA	Summer, Fall
Day of the Week Student Submitted Admissions Application	IR	CA	All days Monday through Sunday
Day of the Week Student was Admitted	IR	CA	All days Monday through Sunday
Month Student Submitted Application for Admission	IR	CA	All months January through December
Month Student was Admitted	IR	CA	All months January through December

Admission Turnaround Time (Days)	IR	NU	Application admitted date subtracted from the application submitted date equals the number of days it took the student to become admitted after applying
Timing of Application from Beginning of Semester (Days)	IR	NU	Date of first day of classes subtracted from the application submitted date equals the number of days the application was submitted before the start of the semester

Each of these individual data sets from multiple sources and reports required matching criteria to combine the information into one master report. All data sets were reviewed by multiple institutional stakeholders to confirm accuracy and consistency across the data. A unique student identification number was used to match the information from one report to another. Each report was cleaned for missing or incomplete data and then combined into one master data set. The data cleaning process is described in the procedure section that follows.

Procedure

The first research question is structured to identify which factors influence online student enrollment decisions. Those factors are then used in a binary classification problem to develop a model capable of predicting and categorizing students into one of two categories: 1) accepts admissions offer and enrolls and 2) does not enroll. When presented with a newly admitted student, the goal is to have the model correctly predict if that applicant will enroll or not based on the student's application data. This procedure is known as supervised machine learning, in which data from previous admissions cycles when the enrollment outcome was known was used to train the data to predict future enrollment cycles.

To begin this inquiry, cleaning and processing the data was a critical step in organizing the information and preparing it to be used in the predictive model. The majority of the data cleaning process was completed during the series of collaborative meetings across multiple WSU departments and resulted in 4620 unique applications. However, additional data cleaning was necessary after the reports were generated to a desired level of satisfaction, which limited the final population to 3,843 unique applications. For example, the original data sources included international applicants. International applicants are Global Campus applicants, but they work with the Office of International Programs through the admissions process. Because the local context of this research has very little influence on the admissions experience for this applicant population, international applicants were removed from the data set and only domestic applicants were included. Similarly, there are three-degree programs that belong under the umbrella of Global Campus but operate outside of the sphere of influence within the local context. Applicants to those degree programs (electrical engineering, mechanical engineering, and Cesar Ritz HMB) were removed from the data set.

Additionally, when combining data sets from different sources using matching criteria for verification purposes, a population of applicants was identified on the IR report but not the CRM report. After investigating, it was determined that they all were examples of applicants who applied to a different campus but eventually changed campuses to Global Campus. These students were removed from the model because future applications of a model to predict decisions of active applicants will not be able to consider applicants who applied to other campuses because we will not be able to know which students will eventually change campuses at the time of application. For example, when building a predictive model looking at current admitted Global Campus students, we will not be able to identify if admitted students from other campuses will later change their campus to Global Campus. Because we cannot identify them before they change campuses, they will not exist if this model is used to predict live data with current applicants and, therefore are not helpful in the model. These applicants were removed from the data set to maintain consistency in the population evaluated in the predictive model.

Another data cleansing challenge is determining what to do with missing or incomplete data. In this collection of data, it was identified early on that different student types resulted in different GPA categories. For example, a student who had not attended college previously might only have a high school GPA to base an admissions decision on, leaving a college GPA as a missing data item. Alternatively, a student who has previously attended four colleges/universities would have several transfer GPAs to base an admissions decision on but was likely missing a high school GPA. As a result, a *Converted GPA* was created. The converted GPA used the GPA available to make an admissions decision. Therefore, first-year students used a high school GPA, and transfer students used a cumulative college GPA. Creating a converted GPA category was necessary for consistency and eliminated the issue of missing data for this particular variable. Because admissions officers use the GPA category as a way to measure past academic performance regardless of if it was high school or college related, the converted GPA is considered equivalent for this category.

When numerical variables were missing data, median imputation was used to fill in the missing data (Kang H, 2013; Khan S & Hogue, A, 2020; Mallikharjuna et al., 2023). Imputation is a process that replaces the missing data with the median of all available data to limit the impact of limited missing information. In this data set, the only form of imputation that was utilized was on the converted GPA category. It is worthwhile to note that researchers consider median imputation to be a reasonable estimate to substitute missing data from a normal distribution (Kang H, 2013; Khan S & Hogue, A, 2020). This imputation method was also used by Basu et al. (2019) to replace the missing GPA values in their study, so the method has precedence. However, median imputation may lead to inconsistent bias and an underestimation of errors. Using this method allowed for the continued use of the GPA variable, however, provides a limitation to the variable worth noting.

Non-Degree seeking and post-baccalaureate applicants are not required to provide an official GPA on their application, therefore, the data was missing from 424 applicants. I used median imputation to replace the missing values with the median GPA from all applicants to mitigate the effect of any outliers. In the data analysis chapter, I evaluated the converted GPA variable including and excluding non-degree applicants to measure the impact of utilizing median imputation on the variable and found no meaningful change.

Every categorical variable that was missing in more than 10% of the cases was completely removed from the data set. For example, the original data set provided information on an applicant's legacy status, or an output that indicated whether a student's parents had previously attended the university. During the data cleaning phase, it was discovered that legacy data was only collected on first-year students, equating to approximately 85% of the variable data missing. This variable was removed for the purpose of this study due to the large portion of missing data. The remaining categorical data was coded based on the available categorical options for each variable. Finally, all the binary variables were transformed into dummy variables prior to cleaning the data to simplify the analysis phase. The binary variables were all complete and did not require additional cleaning.

Other variables were created using the data from the source reports to extract new variables that were potentially more useful in a different format. This process is called features engineering and utilizes domain knowledge to construct features that improve a model's accuracy. For example, my domain knowledge of the behaviors of online adult learners suggests that the earlier an applicant applies, the less likely they are to enroll. This is typically true because early applicants often apply to multiple schools and have more options. As a result, I created the feature *Timing of Application from Beginning of Semester (Days)* to incorporate this domain knowledge because I suspected that the timing of an application may be related to enrollment decisions. This feature was derived from the *application submitted date* variable and subtracted from the known historical first day of classes. Using domain knowledge of the cyclical nature of the admissions cycle, I hypothesized that it was more likely that the timing of an application being submitted is more useful than just the application submitted date by itself, and features engineering allowed to test this hypothesis. The same was true for *days between 'application submitted' and 'admitted' date* among many of the other application timing variables. The speed and which an application is processed is a variable worth investigating, so these features were engineered as new variables from existing ones from the original report.

Another recommendation provided by Basu et al. (2019) was to incorporate geocoding, or a feature that identifies the relative distance between an applicant's geographic location and the location of the university. Research suggests students tend to enroll in institutions that are geographically closer to their homes (Magda et al., 2020; Mattern & Wyatt, 2009). While WSU Global Campus primarily serves online students, the student's zip code was inputted into a distance calculator to determine their relative distance in miles to the zip code of our flagship campus. This is similar to features engineering but is focused on translating a student's zip code into a numerical feature that may result in a stronger predictive ability. In doing this, there were 24 addresses that belonged to mailing facilities at military bases, which did not represent the accurate location of the applicant. For these 24 applicants, I used a process called cold deck imputation. Cold deck imputation is a process of systematically choosing values from an individual who has similar values on other variables (Haukoos, J. & Newgard, C., 2007). For these applicants, I selected and imputed the zip code of their last school attended to eliminate the missing data.

It is important to note though this group was small (n=24), cold deck imputation often underestimates variance because the variable used was selected only from the list of existing variables within the data set (Haukoos, J. & Newgard, C., 2007). The large data set helps minimize the impact on the variance, however, imputation always includes a level of bias that creates a noteworthy limitation.

Finally, logistic regression requires a reference variable to compute comparisons for categorical variables. The final cleaning step involved recoding categorical variables to identify the reference variables in the model. The normative category, or the most frequent category, was selected as the reference variable for all categorical variables.

Data Analysis

At this stage, the data has been collected, cleaned, and prepared to build a predictive model of student enrollment decisions. First, the data was randomly divided into training and testing sets, where 80% of the data was assigned as a training set and the remainder was reserved for testing. Before applying the training data to the machine learning model, I used features selection to identify and isolate the variables with the most predictive power and answer the first research question. Features selection is a process that removes redundant or irrelevant data to improve the accuracy and efficiency of a machine learning algorithm (Grus, 2015). The goal is to identify which variables do not significantly add value to the model and which variables are highly correlated and therefore are redundant. There are many reasons for eliminating variables in a model. Some variables can simply be too complex to adequately implement into the model. Others, like application terms, are irrelevant to an applicant's enrollment decision. Additionally, distinguishing between statistical significance and practical significance can eliminate irrelevant data and improve validity (Smart, 2005). In logistic regression, odds-ratios can be used to determine if a statistically significant variable has practical significance. The odds-ratios measure how many times larger the odds of an outcome occurring is for one unit of an independent variable. In essence, it provides a numerical value that displays the direction and strength of a relationship between an independent variable and the outcome of a dependent variable. Substantively, statistically significant variables were critically examined for their practical value in the analysis and discussion chapters. This perspective prioritizes practical value as a way to select the features among statistically significant variables. Variables that were statistically significant but had very small effect sizes were considered for removal based on their practical value. After completing the process of features selection, the original 24 variables were narrowed down to 10 final variables, which are described in the results chapter.

Finally, previous researchers have tested several well-known machine learning techniques on questions of student enrollment decisions, including logistic regression, decision trees, SVM, naïve bayes, 10-nearest neighbors, random forests, and gradient boosting, among others (Chang, 2006; & Basu et al., 2019). The conclusions continue to identify logistic regression as the machine learning technique that most accurately predicts student enrollment decisions. Logistic regression is utilized to identify the model with the best fit that describes the relationship between a series of independent variables (IVs) and a binary dependent variable (DV) by producing a 'probabilistic value' that one of the binary DVs will occur (Millar, 2011).

Since testing multiple machine learning models was outside the scope of this research, this research focused on optimizing the model's performance and used logistic regression as the machine learning algorithm to answer the second research question.

Phase 2: Research Questions 3 & 4

Phase two was designed to answer the following research questions:

RQ3: How do university enrollment professionals interpret the factors that influence online student enrollment decisions and the factors that do not?

RQ4: How does the knowledge of factors that influence online student enrollment decisions affect enrollment professionals' perspectives, decision making, and strategy development?

Participants

The participants of phase two included seven enrollment professionals within the local context who work directly on formulating and implementing enrollment management strategies. Participants included three student-facing admissions counselors, two mid-level enrollment managers from marketing, enrollment management, and admissions, and two senior-level division leaders. This sample of participants intentionally ranged in years of experience from 0-21+, expertise from admissions, student services, and marketing, and various levels of seniority in order to provide a cross-section of enrollment management professionals with unique perspectives. Participants were 71.4% female and primarily described themselves as white/Caucasian. Seven participants were recruited to participate, and all were familiar with and directly influence the work of admissions and recruitment at WSU Global Campus.

Research Design

Phase two continued step five of the CRISP-DM framework and began the work of step six, engaging participants in planning, discussion, and deployment. Specifically, participants completed an anonymous survey that asked them to select and rank the top ten independent variables identified during the previous phase to the degree to which each had the greatest or least impact on the enrollment outcome of an admitted applicant. Additionally, participants rated each of the 24 variables to the level they believed the variable impacted student enrollment outcomes on a Likert scale from one to six. After the survey was completed, participants attended a workshop to discuss the findings of the model and compare the model with their own perceptions of features that influenced enrollment decisions. This innovation was designed and guided by the four-stage theory of organizational change framework presented in chapter two. Participants were provided with a summary of the model and survey results to review prior to attending the workshop (Appendix B). The workshop was designed to facilitate conversation about both sets of findings and discuss where participants' answers aligned or deviated from the empirical data. The workshop concluded after a final discussion of how this information could be utilized to inform enrollment strategy and decision-making. Finally, a post-workshop survey was completed one week after the workshop to measure the impact of the workshop and collect data about the participants' perspectives.

Instruments & Data Sources

Participants completed a 40-question pre-workshop survey conducted online using Qualtrics (Appendix A) one week before the workshop. The survey was designed first to capture work-related demographic information such as the total years of higher education experience, the number of years working specifically in enrollment management, and the portion of their career spent working directly or indirectly with students. Then, participants ranked each of the independent variables based on the level of influence they believed the variable has on the enrollment outcome of admitted students on a six-point Likert scale ranging from *no influence* to *strong influence*. Using a six-point Likert scale removed neutral or unknown options to obtain responses to each question and eliminate the uncertainty presented when neutral options are permitted (Reips, 2010). Participants did have the ability to select "prefer not to answer" if they wished. Finally, participants selected and ranked the top ten variables in order that they believed most influenced student enrollment decisions. Each of the ranking questions was coded on a scale from one to six to indicate if the participant thought the variable had *no influence* to *strong influence*. Ranking questions were designed and written in the affirmative to maintain consistency across all questions.

Only omitting the demographic questions, the post-workshop survey was an exact duplicate of the pre-workshop survey. However, an additional set of open-ended questions at the end of the survey collected additional qualitative data to discover if and how the model and workshop influenced their professional perspectives and practice. The post-workshop survey was issued one week after the workshop to measure changes in participant perspectives. During the workshop, participants were asked a series of semi-structured questions to help facilitate conversation about the findings of the survey, the model, and how that information could be used to influence the enrollment work the participants are responsible for. The participants' responses to the workshop discussion produced qualitative data collected in the form of observational notes and a transcript that are explored in the data analysis section.

Procedure

All participants were sent a recruitment letter via email asking them to participate in the study. The participants who agreed to be in the study were then emailed a consent form. When the participants signed and completed the consent form, they were automatically routed to the pre-workshop survey and asked to complete the survey within a two-week deadline. The workshop was scheduled on the participants' calendars one week after the survey deadline. Research participants were instructed to provide a unique identification code that would protect their identity but allow for the researcher to match survey submissions for later analysis. Once all surveys were collected, a quick analysis of the survey submissions was conducted, and a summary of the survey results compared to a summary of model's findings was produced in a one-page informational sheet (Appendix B). The one-page results summary was emailed to all participants three days before the workshop, requesting that participants review the information sheet and take notes on their observations to prepare for the workshop.

The audio-recorded workshop was scheduled for 1.5 hours and was facilitated to follow the four-stage theory of organizational change framework. The agenda can be found in Appendix C. The first 10 minutes included an overview of the workshop and presented the problem of practice. I verbally overviewed the model and survey findings for the next 20 minutes. Finally, the first half of the remaining hour was dedicated to discussing the findings and facilitating a conversation about the findings. Participants were asked to discuss the results of the model and what they found interesting, surprising, validating, or challenging to their pre-model expectations. Additionally, participants were asked to reflect on the results of the pre-survey, how those results compared to the model and how their perspectives were similar or different from those of their enrollment peers.

The last half hour was dedicated to discussing how the findings could be used in their professional practices. Participants were asked to generate practical suggestions on how the model and the knowledge produced by the model could be used to generate strategy and achieve shared enrollment goals. Evaluating the model's utility and practicality were central to the semi-structured conversation prompts to elicit responses that would answer the research questions about how the model affects enrollment professionals' perspectives, decision-making, and strategy development. At the conclusion of this workshop, this section was primarily focused on evaluating the merits of using or not using the new information.

I played a participant observer role, led the discussion, and took notes on participant contributions. The workshop was recorded for future transcription and analysis, and the recording was stored on a password-protected computer. Finally, the post-workshop survey was issued one week after the workshop, and participants were asked to complete the survey within a two-week deadline. The results of both surveys, the data from the workshop, as well as the outcome of the model were later integrated to provide a fuller description to answer the research questions.

Data Analysis

Pre-/post-survey data were analyzed independently using descriptive statistics to get a general sense of the response data. Then, the results of both surveys were compared using a t-test for dependent samples to answer how and to what extent participants' responses changed between their pre- and post-workshop surveys. Dependent sample t-tests were used to compare the mean scores of a single group in different measurements.

The recorded workshop produced qualitative data in the form of observational notes and a recorded transcript which were collectively analyzed in an iterative manner. Qualitative analysis in action research often occurs immediately with the researchers' experiences in the data collection process (Herr & Anderson, 2005). The open-ended questions on the post-workshop survey data produced additional qualitative data in the form of narrative responses. I used an inductive approach to qualitative data analysis and followed Creswell's (2009) steps for analyzing qualitative data. The recorded workshop data was transcribed using an electronic transcription service. The raw transcript was then organized, validated with the recorded audio file, and prepared for data analysis. In addition to the live workshop, I re-listened to the full recorded workshop two additional times during the transcript validation process which assisted in gaining familiarity and a general sense of the collected data. Then, I began the process of coding the transcription and narrative survey responses and maintained a codebook that provided notes and my interpretations and descriptions of the meaning of each code. I completed a full coding session of the qualitative data on three separate occasions until I judged that saturation with the data had been reached and no additional codes were found. At this point, I reviewed the codes produced and organized them into related themes to interpret the meaning of the findings and presented them in a visual table represented in Table 9 and Appendix E. Finally, these findings and a discussion of how the findings were interpreted are reported in the results chapter.

Integration

The final stage involved the integration of quantitative and qualitative results to produce meta-inferences. Meta-inferences are defined as “a conclusion generated through an integration of the inferences that have been obtained from the results of the QUAL and QUAN strands of a MM study” (Teddlie & Tashakkori, 2009, p. 152). While phase one was primarily quantitative, QUAL and QUAN data collection and analysis methods were integrated throughout phase two. In the final step, the mixed methods result of the previous phases were integrated into an interpretation stage and discussed together. According to Johnson and Turner (2003), data integration is a fundamental principle of mixed methods research that results in quality meta-inferences by focusing on complementary strengths rather than nonoverlapping weaknesses. Integration is a critical stage of mixed methods research and is the focus of the final stage.

CHAPTER 4

ANALYSIS AND RESULTS

The purpose of this action research study was to build a predictive model of online student enrollment decisions and evaluate the impact on the perceptions from a team of key enrollment managers within the local context, the online campus of a public research university. The goal was to create a better understanding of my local context within my sphere of influence and explore a gap in the available literature outlined in chapter two. The study was guided by four research questions (repeated below) and was designed in multiple phases. This chapter will present the results of data collection and analysis of each phase as prescribed in the methodology chapter.

Phase One

The first research question explored in phase one used data mining to understand which factors and to what extent do those factors influence the enrollment decisions of online students admitted to a public research university. The second research question used machine learning to understand how and to what extent do those factors predict the enrollment decisions of online students admitted to a public research university. To answer these research questions, as described in Chapter 3, I worked with multiple departments across the local context to collect, compile, and clean application data from admitted undergraduate students who applied for the fall semesters spanning three years from 2020 – 2022. This process resulted in 24 independent variables that were used to answer the first two research questions using features selection and logistic regression as methods of analysis. These features were used to categorize each applicant into a binary classification problem aimed at predicting if an admitted applicant will enroll or will they not enroll.

A total of 3,843 applicants were included in the study. Of this population, 57.1% became enrolled students while the remaining did not enroll. Female students outnumbered male students 69.8% to 30.2%, respectively, while the average age of the entire population was slightly over 30 years old when they applied. 35.8% of applicants reported they were first-generation students while 83% reported they were pursuing their first bachelor's degree. The population was comprised of primarily white (61.9%), Washington residents (76.7%), transfer students (60.2%), who transferred an average of 51.6 semester credits.

Table 3

Descriptive Statistics - Continuous Variables (n=3,843)

	Range	Min	Max	Mean	Median	SD
Age	57.4	16.3	73.7	30.2	27.6	9.8
Converted GPA	2.7	1.3	4	3.1	3.1	0.5
Mailing Address Distance from Pullman Campus	5893.3	0	5893.3	440.7	229.5	606.9
Admission Turnaround Time (days)	340	0	340	34.8	20	42.3
Timing of Application from Beginning of semester (Days)	376	0	376	143.6	141	78.7
Total Transfer Credit	315.6	0	315.6	51.6	57.6	41.7

Table 4

Descriptive Statistics - Categorical Variables

	N	Percent
Enrolled Indicator		
Did Not Enroll	1647	42.9%
Enrolled	2196	57.1%
Total	3843	100
Gender		
Female	2682	69.8%
Male	1161	30.2%
Total	3843	100
First-generation		
No	2467	64.2%
Yes	1376	35.8%
Total	3843	100
Military Affiliation		
No	3559	92.6%
Yes	284	7.4%
Total	3843	100
Student Type		
First Year	405	10.5%
Former Student Returning	734	19.1%
Non-Degree Seeking	389	10.1%
Transfer	2315	60.2%
Total	3843	100
Degree Type		
First Undergraduate Degree	3191	83.0%
Other/Non-Degree Seeking	416	10.8%
Post-Baccalaureate	236	6.1%
Total	3843	100
Washington Resident		
Non-Resident	895	23.3%
Resident	2948	76.7%
Total	3843	100
Ethnicity		
American Indian/Alaska Native	47	1.2%
Asian	247	6.4%

Black/African American	190	4.9%
Hispanic/Latino	584	15.2%
Native Hawaiian/Other Pac Island	33	0.9%
Not Reported	100	2.6%
Two or More Races	265	6.9%
White	2377	61.9%
Total	3843	100
Application Semester		
Fall	2589	67.4%
Summer	1254	32.6%
Total	3843	100
Application Month		
January	509	13.2%
February	392	10.2%
March	529	13.8%
April	488	12.7%
May	472	12.3%
June	428	11.1%
July	529	13.8%
August	76	2.0%
September	51	1.3%
October	83	2.2%
November	121	3.1%
December	165	4.3%
Total	3843	100
Admitted Month		
January	210	5.5%
February	371	9.7%
March	459	11.9%
April	610	15.9%
May	524	13.6%
June	422	11.0%
July	634	16.5%
August	436	11.3%
September	21	0.5%
October	37	1.0%
November	40	1.0%
December	79	2.1%
Total	3843	100

Applied Day of the Week		
Monday	656	17.1%
Tuesday	704	18.3%
Wednesday	631	16.4%
Thursday	637	16.6%
Friday	582	15.1%
Saturday	304	7.9%
Sunday	329	8.6%
Total	3843	100
Admitted Day of the Week		
Monday	639	16.6%
Tuesday	767	20.0%
Wednesday	787	20.5%
Thursday	726	18.9%
Friday	697	18.1%
Saturday	128	3.3%
Sunday	99	2.6%
Total	3843	100

Results

Logistic regression was used to analyze the relationship between 24 independent variables on the enrollment outcome of a group of admitted students (dichotomous dependent variable). Three models were run with the training data set using forward, backward, and enter methods of logistic regression. The findings from the three preliminary models, coupled with practitioner knowledge, contributed to the final fourth model evaluated with a testing data set. Each of the models identified different variables and was trained by the proceeding models.

Model one used the enter method and included all independent variables into a single model. The model looked at all of the variables simultaneously and had a run time of approximately 30 minutes to produce the results.

Model one was statistically significant ($\chi2(1401) = 2319.982$, $p < .001$). The model explained 60.9% (Nagelkerke R2) of the variance in enrollment outcome and correctly classified 80.8% of cases. This model identified eight statistically significant predictor variables ($p < 0.05$) summarized in Table 7. While this method produced a statistically significant model with a strong classification/prediction accuracy, it also included all variables which increased the complexity and run-time of the model. To minimize run-time and eliminate irrelevant variables, stepwise regression using the forward and backward methods produced the next two models.

Model two used the backward stepwise method of logistic regression. This method begins with the first step that matches model one, including all variables available. Then the model identifies and removes the least significant variable, one at a time, in each of the subsequent steps. The model performs this process repeatedly, removing one additional variable in each step, until the default elimination criterion has been reached. The backward stepwise method of logistic regression is a simple way to identify the least significant predictor variables and remove them from subsequent models.

SPSS applies a default elimination criterion of a p-value > 0.1. This method ran 11 elimination steps, resulting in a statistically significant model ($\chi2(1378) = 2294.937$, $p < .001$). Model two explained 60.4% (Nagelkerke R2) of the variance in enrollment outcome and correctly classified 80.5% of cases. These results show that the removal of 10 variables had very little impact on the model capability. Over the 11 elimination steps, model two's Nagelkerke R2 decreased only 0.5% and the classification accuracy decreased only 0.3%, indicating that the variables removed had little to no impact on the predictive ability of the final model. Table 5 chronologically illustrates each step, the variable removed, the impact on correct classification, and the goodness of fit for the newest iteration of each model. Each subsequent removal step excludes all of the variables above that step. An interesting observation is that the model's classification showed improved and peaked at 81% correct classification through step seven with nearly zero impact on the goodness of fit. The removal of the last four variables showed only slight decreases in the model's performance.

Model two was more accurate at predicting the *enrolled in classes* category correctly than the *did not enroll* group. By the 11th step, the model correctly predicted the enrollment outcome of applicants who enrolled 89% of the time, while correctly predicting the enrollment outcome of applicants who did not enroll only 69.1% of the time. While type II error was low, type I error, or false positives, were slightly higher. The actual population of applicants enrolled at 57.95% while 42.05% did not enroll. Therefore, model two showed a 31.05% improvement in accuracy at predicting applicants who will enroll and a 27.05% improvement in predicting applicants who will not enroll, then relying on historical yield rates alone.

Table 5

Backward Stepwise Method - Variables Removed

Step	Variable Removed	df	Nagelkerke R Square	Classification Table - Overall Percent Correct
1	Full Model	1401	0.609	80.8%
2	Application Semester	1400	0.609	80.8%
3	Zip Code	1399	0.609	80.7%
4	Applied Day of the Week	1393	0.608	80.7%
5	Distance (miles) from Campus	1392	0.608	80.7%
6	Age	1391	0.608	80.9%
7	Military Affiliation	1390	0.608	81.0%
8	Converted GPA	1389	0.607	80.8%
9	Timing (Days) of application from beginning of semester	1388	0.607	80.9%
10	Ethnicity	1381	0.605	80.4%
11	Student Type	1378	0.604	80.5%

At this stage, model one identified eight statistically significant predictors among the complete list of independent variables. Then, model two removed 10 factors with very minimal impact on the model's utility. However, the backward stepwise method of logistic regression removes a single variable one at a time and then re-runs the model repeatedly until there are no remaining variables that meet the removal criteria. As a result, the runtime exceeded 2 hours, limiting the accessibility of the model. To address the length of time it takes to produce the model, and to identify the top predictive factors, the third model used the forward stepwise method of logistic regression.

The forward stepwise method of logistic regression begins with a null model, which contains no variables. Then, the model runs similarly to model two but in reverse order. In each step, model three added the next most significant variable with a default entry criteria of a p-value < 0.05. This process continues until no remaining variables meet the default entry criteria. Therefore, the forward stepwise method of logistic regression is helpful to identify the most statistically significant predictors within a model.

Model three ran 11 addition steps that resulted in a statistically significant model $\chi2(61) = 722.539$, $p < .001$. Model three explained 23% (Nagelkerke R2) of the variance in enrollment outcome and correctly classified 69.2% of cases. This method shows that the systematic addition of the most significant variables had a much larger impact on the model's capability as compared to the previous models. Compared to model two, model three's goodness of fit decreased by 37.4% and the classification accuracy decreased by 11.3% indicating that this approach had a large impact on the predictive ability.

Table 6 chronologically illustrates each step, the variable added, the impact on correct classification, and the goodness of fit for the newest iteration of each subsequent model. An interesting observation is that the first step with only one predictor accurately classified the enrollment outcome of each applicant 63.9% of the time, while the addition of the next 10 top predictors only improved the classification accuracy by 5.3%. The goodness of fit as represented by Nagelkerke R2 increased only 13.3% from step 1 to step 11. Furthermore, steps within this model show that adding a new variable did not have a meaningful or positive impact on the classification accuracy. The addition of *application month* in step eight had zero impact on classification accuracy while the addition of *admitted day of the week* in step five and *first-generation flag* in step nine decreased the classification accuracy. These are examples of variables that are statistically significant predictors in the model, but the impact on the improvement of the model's performance was unclear.

Similar to model two, model three was more accurate at predicting the *enrolled in classes* category correctly than the *did not enroll* category. By the 11th step, the model correctly predicted the enrollment outcome of applicants who enrolled 83.7% of the time, while correctly predicting the enrollment outcome of applicants who did not enroll only 49.8% of the time. While type II error remained low, type I error, or false positives, were almost equivalent to a coin flip. Model three displayed a 25.75% improvement in predicting the enrollment outcome of an admitted applicant while only a 7.75% improvement in predicting applicants who will not enroll than if we just relied on the historical yield rates alone.

Table 6

Forward Stepwise Method - Variables Added

Step	Variable Added	df	Nagelkerke R Square	Classification Table - Overall Percent Correct
1	Student Type***	3	0.097	63.9%
2	Washington Residency***	4	0.129	65.1%
3	Admitted Month***	15	0.169	67.1%
4	Degree Name***	38	0.196	68.4%
5	Admitted Day of the Week***	44	0.207	68.0%
6	Degree Type***	46	0.213	68.5%
7	Total Transfer Credit***	47	0.217	68.9%
8	Application Month***	58	0.225	68.9%
9	First-generation Flag**	59	0.227	68.7%
10	Distance (miles) from Campus*	60	0.228	69.0%
11	Timing (Days) of Application Complete*	61	0.230	69.2%

$*p < .05$, $**p < .01$, $***p < .001$

Stepwise methods of logistic regression provide a replicable and impartial way to decrease the quantity of factors as compared to manually selecting variables based on practitioner knowledge, which can introduce personal biases. However, there can also be value in applying practitioner knowledge to include variables that may not be statistically significant predictors or eliminate significant predictors that do not have practical value. The process of automated features selection is designed to support, not replace, practitioners' experience. While phase two of the research is intended to explore the practitioner perspectives, I attempted a fourth model of logistic regression using the testing data set to combine common factors across the previous three models from the training data set to explore the outcome of merging statistical and practical significance informed by empirical data and practitioner knowledge.

Table 7 illustrates the factors from model training that were retained or eliminated in each model, and the factors selected in the final fourth model. First, I identified which factors I would exclude from the model. I reviewed the list of the factors removed during the backward stepwise method of logistic regression (model two) which systematically identified and removed the least valuable predictors from model two. I then compared those predictors to the ones excluded in model three. I found eight matches that were not included in either model and I removed those from the original list of 24 factors.

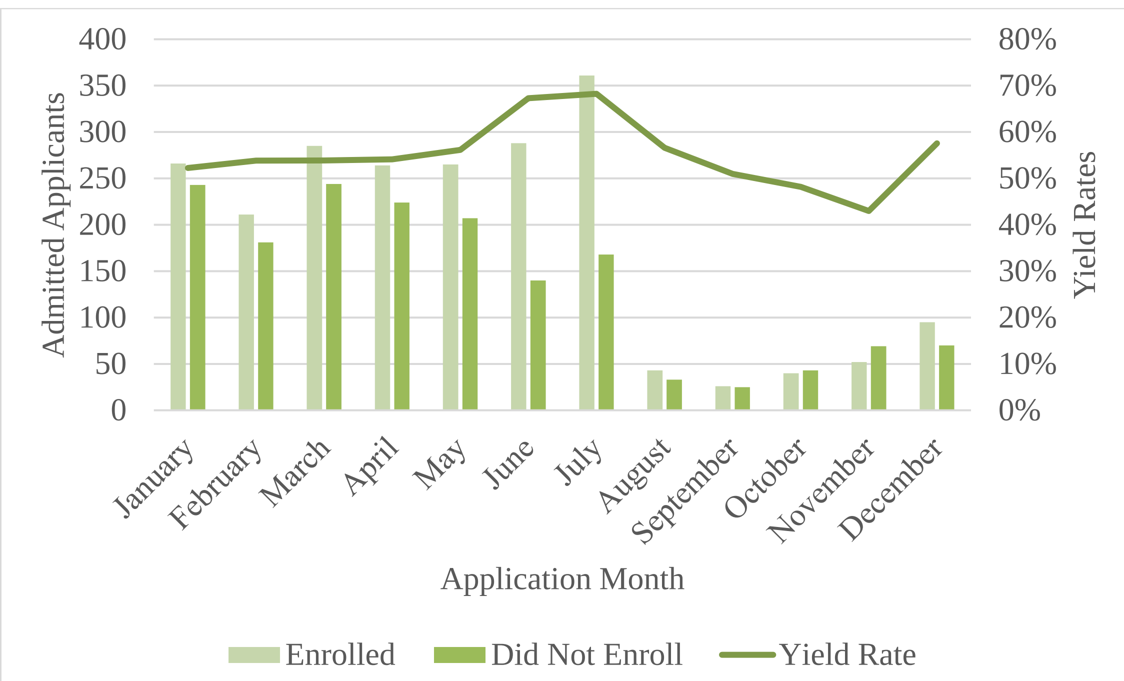
Of the remaining 16 factors, I reviewed the most statistically significant factors from all three models. Seven common factors were statistically significant across all of the models (illustrated in Table 7). I selected those commonalities as my first seven factors in the fourth model. Then, I reviewed each of the remaining nine factors that were statistically significant in at least one of the previous models and determined to keep or eliminate them from model four. I included the *admissions turnaround time (days)*, *application month*, and *student type* variables in the final model. Factors with a small number of categorical values, such as those in student type and day of admission, could lead to an over interpretation of the sign of estimates. While I chose to include these factors in the model, it is a noteworthy limitation to these types of variables in this type of study. *Gender* and location-based factors including *state*, *county, city, distance from flagship campus (miles),* and *last school attended* were removed due to their complexity, low odds ratios, and potential to introduce bias into the model.

While it was only statistically significant ($p < .05$) in the forward method implemented in model three, I chose to retain the *admission turnaround time (days)* factor in the final model because the classification accuracy improved when the factor was introduced to that model. The odds ratio of -0.004 is relatively small, but it is important to note the odds ratio is the change in odds of an enrollment outcome based on a single unit change of the variable. In this case, the odds of a student enrolling decrease 0.4% for every additional day before an admissions decision. Alone, this has limited meaning to enrollment managers. However, interpretation of this variable becomes more meaningful when extrapolated across multiple days to translate the outcome to more of an operational scale. For example, for every 10 days an applicant is waiting for an admissions decision, the odds they will enroll decrease by 4%. The coefficient and standard error adjust to -0.04 and 0.02, respectively, when reducing the turnaround time by 10 days. For context the mean admission turnaround time is 34.8 days, meaning changes to this variable have a more meaningful interpretation at operational scale. Extrapolating the odds ratio across a different unit of measurement makes this variable and others more practical, interesting, and potentially useful to enrollment managers. Additionally, it is presumed that students who receive an admissions decision quickly have a better experience than students who must wait longer to receive the same information. I suspect that students with better experiences enroll at a higher rate, especially those applying to multiple institutions. Student experience during the admissions process is not a measurable factor in this study; therefore, I relied on my experience to retain this factor in the final model.

Application month, which was statistically significant in model three ($p < .001$), was another factor I selected to retain based on domain knowledge and experience. Figure 5 illustrates the admissions funnel's cyclical nature using the study's data. The closer a student applies to the application deadline (July), the higher their likelihood of enrolling becomes (yield rate). As a result of the impact of application timing on historical yield rates, I chose to keep this factor in the fourth model.

Figure 5

Enrollment Yield by Application Month



Finally, I chose to keep *student type* in the model. This factor was the most significant predictor ($p < .001$) identified in model three, although it was not found to be significant in the other models. As the first factor introduced in step 1, that factor alone correctly classified the student enrollment outcome 63.9% of the time. Transfer applicants were the most frequent student type (n=2315) and were used as the reference category. While first-year applicants were 26.8% less likely to enroll than transfer applicants ($p < .05$), former students returning were 31.5% more likely to enroll than transfer applicants ($p < .05$). Because of the fairly strong odds ratios and the statistical significance of each category, student type was retained for the final model. While I chose to include this factor in the model, it is a noteworthy reminder that factors with a small number of categorical values, such as transfers being the majority of the student type, could lead to an over interpretation of the sign of estimates. This factor was retained for the final model, but this limitation should be noted.

Model four included 10 of the original 24 variables and was statistically significant $\chi2(60) = 715.886$, $p < .001$. The model explained 22.8% (Nagelkerke R2) of the variance in enrollment outcome and correctly classified 69.2% of cases. The model performed similarly to the forward method due to the alignment of the variables selected with both the training and testing data sets. The runtime was nearly instantaneous due to using the enter method of logistic regression and the removal of complex categorical variables like those associated with an applicant's geographic location. Compared to the 1401 degrees of freedom in the first model, the fourth model contained only 60. Model four continued to maintain little type II error (83.7% correct classification) but was less effective in predicting applicants that did not enroll (49.8% correct). While model 4 did not have the highest classification accuracy from all the models, it represented the factors that were most influential in predicting the enrollment outcome of admitted applicants and answered the research questions. The goal of phase one was to narrow the factors down to identify the most and least significant variables, then discuss them in phase two with enrollment professionals. For this reason, model 4 was the model presented to participants in phase two of the research where participants discussed the individual factors in detail. The variables included and excluded from the final model and all previous models are summarized in Table 7. The odds ratios, confidence intervals, and significance levels of each variable in the final model are illustrated in Appendix D.

Table 7

Summary of Factors from Each Model

Model Method Data Set	M1 Enter Train	M2 Backward Train	M3 Forward Train	M4 Enter Test
Significance	< .001	< .001	< .001	< .001
Chi-Square	2319.982	2294.937	722.539	715.886
Degrees of Freedom	1401	1378	61	60
Classification Accuracy	80.8%	80.5%	69.2%	69.2%
Nagelkerke R2	0.609	0.604	0.230	0.228
Variables:				
Washington (WA) Residency	✓*	✓*	✓***	✓***
Admitted Month	✓*	✓***	✓***	✓***
Application Major (Degree Name)	✓**	✓**	✓***	✓***
Admitted Day of the Week	✓**	✓***	✓***	✓***
Degree Type	✓***	✓***	✓***	✓***
First-generation Status	✓***	✓***	✓**	✓*
Total Transfer Credits	✓**	✓*	✓***	✓***
Admission Turnaround Time (Days)	✓	✓	✓*	✓*
Application Month	✓	✓	✓***	✓***
Gender	✓*	✓*		
Student Type	✓		✓***	✓***
State	✓	✓***		
County	✓	✓***		
City	✓	✓***		
Last School Attended	✓	✓***		
Distance from Flagship Campus (Miles)	✓		✓*	
Zip Code	✓			
Application Semester (fall vs. summer)	✓			
Age	✓			
Ethnic Origin Description	✓			
Cumulative Converted GPA	✓			
Military Affiliation	✓			
Timing of Application from Beginning of Semester (Days)	✓			
Applied Day of the week	✓			

**p < .05, **p < .01, ***p < .001*

Summary – Phase One

Phase one was designed to identify which factors influence the enrollment outcomes of students admitted to an online university and how well those factors predict their enrollment decisions. Logistic regression was used as a supervised machine learning tool to evaluate 24 independent variables to answer the research questions. Three statistically significant models were built to identify the most and least predictive variables. The results of the first three models, coupled with practitioner knowledge and evaluation, contributed to the final fourth model presented to participants in phase two of the research.

The first research question asked, which factors and to what extent do those factors influence the enrollment decisions of online students admitted to a public research university? Table 7 illustrates the statistically significant factors from each model, while Appendix D illustrates the odds ratios and confidence intervals for the factors included in model four. The odds ratio illustrates the extent to which the variable had influence on the enrollment outcome and measures the strength and direction of a relationship between the independent variable and the enrollment outcome. Across all four models the factors that influenced enrollment decisions included *Washington residency, admitted month, major/degree name, admitted day of the week, degree type, first-generation status, total transfer credits, admission turnaround time (day*s), and the *application month*.

The second research question asked, how and to what extent do those factors predict the enrollment decisions of online students admitted to a public research university? At the peak, step seven in model two correctly predicted the enrollment outcome of students 81% of the time, although this model did include non-significant variables. All models performed better at correctly predicting "enrolled" as an outcome than the "not enrolled" category, and all models performed better than the null model (57.95%). The final 10 factors included in the final model presented to research participants in phase two only included significant factors and correctly predicted the enrollment outcome of the entire population 69.2% of the time, improving upon the null model by 11.25%. Therefore, I can conclude that the final 10 factors (*Washington residency, admitted month, major/degree name, admitted day of the week, degree type, first-generation status, total transfer credits, admission turnaround time (days), application month*, and *student type*) contributed to a statistically significant model capable of predicting the enrollment decisions of admitted online undergraduate students.

Phase Two

The research questions presented in phase two include:

RQ3: How do university enrollment professionals interpret the factors that influence online student enrollment decisions and the factors that do not?

RQ4: How does the knowledge of factors that influence online student enrollment decisions affect enrollment professionals' perspectives, decision making, and strategy development?

To answer these research questions, I developed a workshop supported by a pre- and post-workshop survey to evaluate enrollment management professionals' perceptions within the local context. The data included quantitative and qualitative data sources and the findings are explored in this section.

Research participants were identified and recruited from the local context as professionals with enrollment management experience who held roles responsible for working with the student population from phase one. Participants (n=7) ranged in years of higher education experience from less than 1 to more than 21 and held positions from student-facing level roles through senior leadership level roles. There were two male and five female participants who primarily described themselves as white/Caucasian.

After agreeing to participate in the study, research participants completed a pre-workshop survey one week before the workshop. Participants ranked each of the 24 independent variables on a six-point Likert scale based on how much they believed a specific factor influenced the enrollment outcome of an admitted student. Then, participants selected and ordered the top 10 variables they believed influenced the dependent variable. Finally, the pre-workshop survey included an optional open-ended textbox that asked if there were factors not included in this study that they believed would influence the enrollment outcome of an admitted student. Six participants indicated yes to that question and provided a narrative response.

A summary of model four that compared the participant responses from the pre-workshop survey (Appendix B) was emailed to participants three days before the workshop. Then, participants participated in the workshop which produced qualitative data analyzed in this section. Finally, participants completed the post-workshop survey one week after the workshop, producing quantitative and qualitative data in the form of open-ended questions. The pre/post-workshop data, and the qualitative data produced by the workshop, are the focus of the next two sections.

Pre-Workshop Survey

The pre-workshop survey was comprised of three major sections. First, participants scored each of the 24 possible variables on a scale from 1-6. Then, participants selected the top 10 variables and ordered them from most to least influential. Finally, participants had an opportunity to respond to an open-ended question asking their suggestions. Because participants completed the pre-workshop survey before they were introduced to the model, participants' responses are based on their own experiences.

Summarized in Table 8 and visually illustrated in Appendix B, participants aligned fairly well with the factors selected in model four. Of the 10 statistically significant factors, participants mostly agreed that *total transfer credits, first-generation status, admitted month,* and *Washington residency* were factors that influenced an applicant's enrollment outcome. Meanwhile, there seemed to be more variance in the perspectives of *degree type, student type, application major*, and *admitted day of the week*.

Table 8

Descriptive Statistics – Pre-Workshop Survey

	N	Min	Max	Mean	SD	Responses Selected in Top 10	In Model 4
Number of Days Applied Before Start of Semester	7	4	6	4.57	0.79	100%	
Total Transfer Credits	7	4	6	5.00	0.82	86%	✓
Month Admitted	7	4	6	4.71	0.95	86%	✓
WA Residency	7	3	6	4.86	1.07	71%	✓
First Gen Status	7	4	6	4.57	0.98	71%	✓
Admission Turnaround Time (Days)	7	4	6	5.00	0.82	57%	✓
GPA	7	2	5	4.14	1.22	57%	
State	7	4	6	4.86	0.90	43%	
Month App Submitted	7	4	6	4.71	0.95	43%	✓
Age	7	3	6	4.57	1.13	43%	
Military Affiliation	7	3	6	4.43	1.13	43%	
Degree Type	7	2	6	4.29	1.38	43%	✓
App Semester (Summer Vs. Fall)	7	2	6	4.29	1.70	43%	
Ethnicity	7	2	6	4.14	1.35	43%	
Student Type	7	2	6	4.14	1.35	43%	✓
Last School Attended	7	2	6	3.43	1.27	29%	
Gender	7	1	4	2.86	1.07	29%	
Zip Code	7	2	6	4.00	1.29	14%	
City	7	2	5	3.71	1.38	14%	
Distance From Campus	7	2	6	3.71	1.38	14%	
Application Major	7	2	5	3.14	1.22	14%	✓
Admitted Day of the Week	7	1	4	2.29	1.25	14%	✓
County	7	2	6	3.71	1.50	0%	
Applied Day of the Week	7	1	4	2.00	1.00	0%	

Of the 14 factors that were not included in model four, 100% of participants ranked the *number of days the application was submitted before the start of the semester* in their top 10 (Mean = 4.57, SD = 0.73). Other non-significant factors that participants ranked above the average score for each variable included an applicant's *state, age*, and *military affiliation*. *Application semester* (summer vs. fall) was the most disagreed upon factor with a standard deviation of 1.7. Participants mostly agreed that geographic-based factors had little influence with some of the lowest mean scores coming from *county, city, zip code*, and *distance from Pullman (miles)*. The lowest ranked factor was the *applied day of the week* with a mean of two.

Workshop

The 90-minute workshop consisted of three primary sections and was recorded/transcribed for data analysis. For the first 30 minutes, I introduced participants to the research study, the data collection methods, the results of the predictive model, and the pre-workshop survey. For the next 30 minutes, I facilitated discussion that allowed research participants to ask questions and reflect on the results of the model compared to the results of the survey. Finally, in the last 30 minutes I asked all participants to answer and discuss three distinct questions outlined in the workshop agenda (Appendix C). After the completion of the workshop, I reviewed, coded, and categorized the transcripts into related themes displayed in Appendix E with examples provided in Table 9.

Table 9

Themes, Codes, & Example Statements

Themes	Codes	Example
1. Desire to Understand	1.1 Reservation	"There are non-predictive factors that are still important that aren't statistically important so I'm not sure you cannot completely rely on the predictive model." (P2-Workshop)
	1.2 Seeking Clarity	"I had a question about one of the nonsignificant factors when you identified the application semester, summer versus Fall, which I thought was interesting. I would have expected fall versus spring. Could you talk a little bit about what that means? Maybe I'm not interpreting it correctly." (P1-Workshop)
	1.3 Wanting to Learn More	"As I was looking at this, I found myself wanting to know more about each category, like the States. Like, what is the yield for New York vs. Connecticut. Or the majors, like, which major is better, and which is worse?" (P6-Workshop)
	1.4 Connecting to Local-Level Knowledge or Experience	"I would assume that, like, former students returning applicants are probably way more likely to confirm than other people, and that probably would have an influence over the data." (P7-Workshop)
	1.5 Acceptance / Understanding	“Yeah, thank you. That makes sense. And I think it could have accounted, perhaps, for the way we answered questions, too, how we interpreted the definition of a variable." (P3-Workshop)

2. Desire for Application	2.1 Optimism	"Just to add, I feel like this information you have really helps us ask the right questions in each category where, like it, kind of directs the conversation where we need to spend more time exploring." (P3-Workshop)
	2.2 Recommendation / Suggestion	"So, I've never built a predictive model, but I read about them and look them up and stuff like that. 10 seems like a lot of predictors. Can we build a less complex model with fewer factors that is just as predictive?" (P3-Workshop)
	2.3 Opportunity to Improve Upon Model	"There are a lot of models out there that predict persistence. But I don't know if I've read anyone that starts at the admissions level, so that would be a very, very interesting way to expand the model." (P5-Workshop)
	2.4 Application to Strategic Planning	"Another thought I had, and this I guess this kind of goes with like our strategic planning goals, when I saw the fact that last school attended was not on the list, I was shocked. But then I also thought like, this is definitely an opportunity for us to learn more about that category for planning purposes." (P4-Workshop)
	2.5 Application to Decision Making	"I want to use the data so we don't exclude groups of people or underserved audiences just because they don't historically yield." (P1-Workshop)

Research question three asked how university enrollment professionals interpret the factors that influence online student enrollment decisions and the factors that do not. I reviewed, coded, and analyzed the workshop transcripts on three separate occasions. During this analysis, I found that research participants initially approached the factors with a sense of curiosity, caution, and a desire to understand the data with which they were presented. For example, participant two said, “I'm weary of the admitted day of the week variable because I don't understand it" while participant four said, “We need to look into the data further and understand what is influencing each of those factors.” These examples and others seemed to display both a sense of uncertainty, caution, and interest early in the workshop and resulted in related codes.

As the workshop progressed, transcript codes slowly began to shift from a sense of reservation and clarity seeking, to tones of curiosity, optimism, opportunity, and application to the local context. For example, participant six added “As I was looking at this, I found myself wanting to know more about each category” and “I wonder if age will change in the future since we are getting a larger portion of traditional-aged students applying.” These examples and others in Appendix E show as participants became more familiar with the data and the model, their contributions gradually transitioned from reservation and clarity seeking to opportunity and application. This transition in codes became clearer with each coding session I completed.

Theme One: Desire to Understand

Initially, there seemed to be a sense of skepticism about the data the participants were presented with. As I reread and listened to the recorded transcript, I determined that the skepticism I initially interpreted was not that at all. The intent behind the participants’ comments seemed to be founded with a sense of curiosity and exploration. I used the code *reservation* (1.1) to describe when a comment expressed concern, however, the intent behind the reservation codes seemed to be expressed not because there was a lack of trust in the model, but because the participants did not fully understand a particular variable. Comments that were coded as *reservation* (1.1) were often followed by group discussion and codes that connected the information to *local-level knowledge or experience* (1.4) and then to either *acceptance/understanding* (1.5) or *wanting to learn more* (1.3) codes.

An example of this was the *admitted day of the week* variable. Three participants expressed reservations about this variable. Comments included "I'm really wondering about *admitted day of the week*. Now I wonder if you could share some insight into that, because it just doesn't make sense to me." And "I'm weary of the admitted day of the week variable because I don't understand it". As I provided information on how this specific variable data was collected from the office of institutional research, participants began to *apply local context knowledge* (1.4) and ask questions *to seek clarity* (1.2). These comments included “Was it the day the staff entered the admissions decision or the day we send them the admissions letter? Or the day they accepted their admissions? Like, how did you decide what day to use?” And "There is a very rare exception that we would actually admit a student on a Saturday or Sunday.” And in response to another participants contribution, “Well, to your point, a lot of stuff in the system has a delay. Like, say, on a Friday. But the process doesn't run until after midnight, and so perhaps the date wouldn't be recorded until Saturday or Sunday". Participants engaged in a back-and-forth discussion as they collectively worked towards helping each other understand using each of their unique experiences and knowledge of the local setting. After clarity seeking and application of local context knowledge, this particular strand of conversation ended in codes of *wanting to learn more* (1.3) with comments that included “So I want to learn more why that variable was significant." And “It seems like we need to ask IR exactly how that date is documented so we can understand this variable more.” These examples contributions illustrate that as participants asked questions to *seek clarity* (1.2), they followed those streams of conversations with statements illustrating they *wanted to learn more* (1.3) especially if the variable was still in question.

The theme that arose from this section was that the participants had a genuine desire to understand. That theme came through five primary codes that all contributed to the participants’ intent to understand the data and materials with which they were presented. Each of the discussion threads in this section ended with *acceptance* (1.5) or a *desire to learn more* (1.3). Often, that was completed through collaborative discussion and *connection to local knowledge* (1.4) which naturally steered the conversation into the third section of the workshop agenda.

Theme Two: Desire for Application

As participants became more familiar with the data, the workshop transitioned from the second to the third agenda section. Simultaneously, the analysis codes also shifted from codes related to understanding to codes related to application. Participants began to layer previous conversations, meetings, and experiences into the context of the data. Examples of this included, "It just made me think of, like, the goals and the conversations we've had with the strategic planning group and how we can use this to shape those conversations." And “I guess this kind of goes with like our strategic planning goals, when I saw the fact that last school attended was not on the list, I was shocked. But then I also thought like, this is definitely an opportunity for us to learn more about that category for planning purposes." These comments illustrate that participants referenced previous meetings, conversations, and goals that aligned how the information could be used in practice. Additionally, these examples and others illustrate that participants experiences and assumptions were challenged with the use of phrases such as “I was shocked”. As their familiarity with the data increased, these example statements and others in Appendix E illustrate the theme change from a desire to understand to a desire for application.

Codes in this section began to layer *opportunities for application* (2.5) & *strategic planning* (2.4) into the conversation, which led participants to also make *recommendations/suggestions* (2.2) and engage in discussions on *model improvements* (2.3). Participants expressed desires and made requests for things to include, or exclude, in future iterations as well as offered alternative applications for the model. Comments such as "Can we use this to focus on certain groups? For example, can we just explore freshmen students and build our recruitment efforts for a single population instead of doing it for everyone?" And "We need to design our systems to be able to be responsive to data. Like legacy data needs to be collected on all students." And "We can use this data to identify gaps and challenges that need addressed, like outreach to underserved audiences." The discussion took the form of modifying the model or modifying organizational structures to support the potential adoption of the innovation. These contributions, comments, and the codes associated with them increased in frequency the further through the workshop we got, further illustrating a movement from understanding to application. This transition from understanding to application was also visible in the way participants contributed to the collective conversation. The example quotes in this paragraph show a transition from questions such as “can we use…” to statements such as “we can use…” and “we need to…”. Questions seeking understanding were replaced with statements illustrating application as the workshop progressed.

The theme that arose from this section was that the participants also shared a desire to apply the information they learned to their work within the local context. That theme came through five primary codes that all contributed to the participants’ interest in discovering how they could use the information. Each of the discussion threads in this section ended with an *application to strategic planning* (2.4) or an *application to decision making* (2.5) code. Tones slide away from expressing *reservation and uncertainty* (1.1) to expressing *optimism* (2.1) and making *recommendations to improve the model* (2.2, 2.3).

The research design and workshop agenda followed the four-stage theory of organizational change as a framework for innovation. Interestingly, the codes and themes extracted from the workshop discussion also seemed to follow this framework. A collective awareness of an existing problem led to a robust discussion of data in which participants strove to understand and make sense of the information with which they were presented. This theme, a desire to understand, was extracted from five primary codes associated with answering research question three. As participants became more familiar with the data, the conversation naturally flowed toward improvements and application to the local context. This theme, a desire for application, was extracted from five primary codes associated with answering research question four.

These two major themes, extracted from the workshop through 10 primary codes, conclude that research participants wanted to learn and apply the new information to improve the work they do within the local context. Comments such as “I feel like this information you have really helps us ask the right questions in each category where, like it, kind of directs the conversation where we need to spend more time exploring" and, “It gives us a starting place to ask deeper questions” are verbal illustrations of the cyclical spirit of action research in practice. These themes demonstrate participants’ desires to learn and apply innovations to improve local practice and serve as starting points for future cycles of research.

Post-Workshop Survey

One week after the conclusion of the workshop, participants were asked to complete a post-workshop survey. The survey mirrored that of the pre-workshop survey with two important differences. Immediately after the workshop, access to the one-page model & pre-survey summary document (Appendix B) was removed for all the participants. This was done to prevent participants from using the resource to answer ranking questions on the post-workshop survey. While the retention of information was not specified in the research questions, the purpose of the post-workshop survey was to measure if participants’ perspectives changed as a result of participating in the workshop. Further, the post-workshop survey contained five additional open-ended questions designed to allow participants to provide qualitative responses to be compared to the quantitative pre/post-survey analysis.

Table 10 displays how participants ranked each of the 24 independent variables on a scale from 1-6 and the change from pre- to post-workshop surveys. The largest increases in mean score came from the *applied day of the week* (+1.8) and *admitted day of the week* (+1.71) variables. The largest decreases in mean score came from the *state* (-1.06), *military affiliation* (-1.03), *age* (-0.97), and *GPA* (-0.94) variables. These changes in mean score from the post-workshop survey better aligned with the final model and illustrate a potential change in participants' perceptions for these specific variables.

Interestingly, two non-significant factors that were not included in the model showed small increases in mean scores. These included the *last school attended* (+0.57) and *gender* (+0.14) variables. Neither of these factors were discussed at much length during the workshop. Additionally, these changes are relatively small due to the small sample of participants. These changes should be perceived with caution due to the minor change and relatively low level of discussion.

While relatively small changes occurred, scores from significant variables that were included in the model but decreased in mean scores included the *admission turnaround time(days)* (-0.80), *month application submitted* (-0.31), *total transfer credits* (-0.2), *first-generation status* (-0.17), *month admitted* (-0.11), and *degree type* (-0.09) factors. Because the survey sample size (n=7) is small, slight changes in the mean score of each factor likely represent little to no change in the actual perception of those specific variables.

Finally, changes in the standard deviations (SD) from the pre- to post-workshop survey illustrate changes in the relative agreement or disagreement among the responses. Again, the participant population is small, so interpreting changes in SD should be approached with caution. However, the largest decreases in SD (indicating increased alignment in participant responses) came from *application semester* (-1.26), *county* (-

0.92), *city* (-0.8), and *distance from Pullman campus* (-0.7) variables. The largest increases in SD (indicating decreased alignment in participants' responses) came from the *age* (+0.35), *total transfer credits* (+0.25), and first-generation *status* (+0.12) variables. In addition to those, *GPA, ethnicity*, and *degree type* had the largest SDs (or highest levels of disagreement among the responses) on the post-workshop survey.

Table 10 summarizes this information for each factor from the pre- to post-workshop survey. The mean scores were measured on a Likert scale from one to six where participants rated the relative level of influence they believed each factor influenced an applicant's enrollment outcome. A score of one indicated a variable had low impact, while a score of 6 indicated that a variable had a strong impact.

Table 10

Descriptive Statistics - Pre/Post-Survey Results

		Pre-Survey		Post-Survey		Difference	
Factor	N	Mean	SD	Mean	SD	Mean	SD
Age	7	4.57	1.13	3.60	1.48	-0.97	0.35
Ethnicity	7	4.14	1.35	3.40	1.24	-0.74	-0.11
First Gen Status	7	4.57	0.98	4.40	1.10	-0.17	0.12
Gender	7	2.86	1.07	3.00	1.00	0.14	-0.07
Military Affiliation	7	4.43	1.13	3.40	0.93	-1.03	-0.20
Application Major	7	3.14	1.22	3.40	0.73	0.26	-0.48
GPA	7	4.14	1.22	3.20	1.21	-0.94	0.00
Degree Type	7	4.29	1.38	4.20	1.21	-0.09	-0.17
Student Type	7	4.14	1.35	4.40	0.93	0.26	-0.41
Total Transfer Credits	7	5.00	0.82	4.80	1.06	-0.20	0.25
Last School Attended	7	3.43	1.27	4.00	1.00	0.57	-0.27
City	7	3.71	1.38	3.00	0.58	-0.71	-0.80
County	7	3.71	1.50	3.00	0.58	-0.71	-0.92
State	7	4.86	0.90	3.80	0.68	-1.06	-0.22
Zip Code	7	4.00	1.29	3.20	0.89	-0.80	-0.40
WA Residency	7	4.86	1.07	5.00	0.58	0.14	-0.49
Distance From Campus	7	3.71	1.38	3.20	0.68	-0.51	-0.70

App Semester (Summer vs. Fall)	7	4.29	1.70	3.60	0.45	-0.69	-1.26
Applied Day of the Week	7	2.00	1.00	3.80	0.89	1.80	-0.11
Admitted Day of the Week	7	2.29	1.25	4.00	0.82	1.71	-0.44
Month App Submitted	7	4.71	0.95	4.40	0.45	-0.31	-0.50
Month Admitted	7	4.71	0.95	4.60	0.45	-0.11	-0.50
Admission Turnaround Time (Days)	7	5.00	0.82	4.20	0.68	-0.80	-0.13
Number of Days Applied Before Start of Semester	7	4.57	0.79	4.20	0.89	-0.37	0.11

After running descriptive statistics, a t-test for dependent samples was conducted to determine if there was a difference between the pre-and post-workshop survey scores for each variable. Because the mean scores were calculated using Likert scale data and the population is small, analyzing the change in mean scores should be interpreted with caution. With that said, participants showed significant change in the way they scored the applied day of the week ($t_6 = 4.03$, $p = .01$) and the admitted day of the week ($t_6 = 3.29$, $p = .02$) variables with large effect sizes, 1.18 and 1.38, respectively. I suspect this change was driven by the number of questions about the significance of the *admitted day of the week* variable during the workshop. While the *applied day of the week variable* was not included in the model, I suspect participants rated both categories higher on the post-survey (1.8 and 1.71 respectively) because of the amount of attention that specific variable received during the workshop. All other variables did not display a significant change. The outcome of the dependent samples t-test for all 24 variables is summarized in Table 11.

Table 11

Dependent Samples T Test

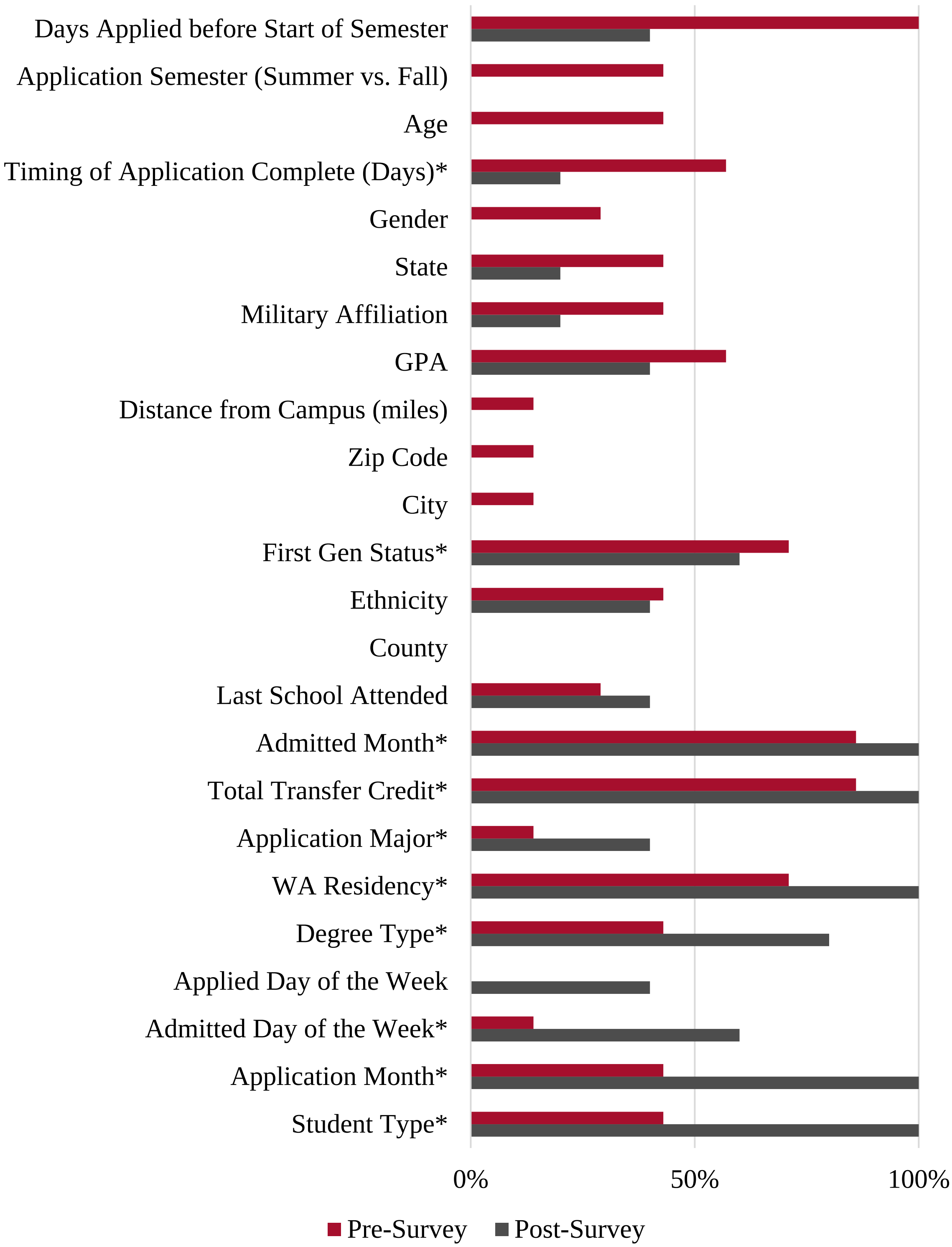
				95% Confidence Interval				Significance
	Mean	SD	Std. Error Mean	Lower	Upper	t	df	Two-Sided p
Age	-0.97	1.10	0.42	-1.99	0.05	-2.33	6.00	0.06
Ethnicity	-0.74	1.39	0.53	-2.03	0.55	-1.41	6.00	0.21
First Gen Status	-0.17	1.42	0.54	-1.49	1.14	-0.32	6.00	0.76
Gender	0.14	1.22	0.46	-0.98	1.27	0.31	6.00	0.77
Military Affiliation	-1.03	1.60	0.60	-2.51	0.45	-1.70	6.00	0.14
Application Major	0.26	1.19	0.45	-0.84	1.36	0.57	6.00	0.59
GPA	-0.94	1.75	0.66	-2.57	0.68	-1.42	6.00	0.21
Degree Type	-0.09	1.60	0.61	-1.57	1.40	-0.14	6.00	0.89
Student Type	0.26	1.32	0.50	-0.96	1.48	0.52	6.00	0.63
Total Transfer Credit	-0.20	1.44	0.54	-1.53	1.13	-0.37	6.00	0.73
Last School Attended	0.57	.90	0.72	-1.19	2.33	0.80	6.00	0.46
City	-0.71	1.80	0.68	-2.38	0.95	-1.05	6.00	0.33
County	-0.71	1.70	0.64	-2.29	0.86	-1.11	6.00	0.31
State	-1.06	1.42	0.54	-2.37	0.25	-1.97	6.00	0.10
Zip Code	-0.80	1.86	0.70	-2.52	0.92	-1.14	6.00	0.30
WA Residency	0.14	1.22	0.46	-0.98	1.27	0.31	6.00	0.77
Distance from Campus (miles)	-0.51	1.40	0.53	-1.81	0.78	-0.97	6.00	0.37

Application Semester (Summer vs. Fall)	-0.69	1.91	0.72	-2.45	1.08	-0.95	6.00	0.38
Applied Day of the Week	1.80	1.18	0.45	0.71	2.89	4.03	6.00	0.01
Admitted Day of the Week	1.71	1.38	0.52	0.44	2.99	3.29	6.00	0.02
Application Month	-0.31	0.71	0.27	-0.97	0.34	-1.17	6.00	0.29
Admitted Month	-0.11	0.84	0.32	-0.89	0.66	-0.36	6.00	0.73
Admission Turnaround Time(Days)	-0.80	1.24	0.47	-1.95	0.35	-1.71	6.00	0.14
Number of Days Applied Before Start of Semester	-0.37	1.07	0.41	-1.36	0.62	-0.92	6.00	0.40

Participants also selected the top 10 factors and ordered them in both survey submissions. Figure 6 illustrates the changes from the pre- to post-workshop survey in the percentage of respondents that selected a factor in the top 10. While all seven participants completed the pre-workshop survey ranking question, only five completed the ranking question on the post-workshop survey. Of the five respondents, the largest increases came from the *student type* (+57%), *application month* (+57%), and the *admitted day of the week* (+46%) variables. Meanwhile, the largest decreases came from the *number of days applied before the start of the semester* (-60%), *application semester* (-43%), and *age* (-37%) variables.

Figure 6

Percentage Selected in Top 10 Variables



**Significant factors included in model four.*

Almost all non-significant variables that were not included in the model showed decreases (12 out of 14) except the *applied day of the week* (+40%) and the *last school attended* (+11%) variable. Similarly, almost all significant variables included in the model showed increases (8 out of 10) except for the *admission turnaround time (days)* (-37%) and the *first-generation status* (-11%) variables. This shows that participants ranked the variables in the post-workshop survey much closer to the model results than the pre-workshop survey rankings, indicating that participants' perspectives may have changed as a result of participating in the workshop.

Finally, the post-workshop survey presented five open-ended questions to collect narrative responses in the form of qualitative data from participants. Participants were asked which factors they were surprised by and why. Every response indicated the admitted day of the week factor was the most surprising. One participant was surprised that ethnicity was not a predictor. Participants were asked, has your perspective changed on which factors influence student enrollment decisions and which do not? Five of the seven responses indicated their perspectives had not changed but provided narrative responses like "I think we can use the data that has come out to fine tune our marketing efforts and to inform strategic priorities" and "it may be useful for marketing and auto generated outreach." Two of the seven responses indicated their perspective had changed and provided narrative responses that included, "It has changed but I do feel left with more questions" and "It is valuable information but is worth further investigation on the results and what influences those factors." Finally, participants were asked, can the results of a model that predicts student enrollment decisions impact enrollment efforts, decision making, and strategy development? All seven participants concluded "yes" and provided a narrative response. These responses include statements such as “it can drive our communication and outreach”, “the data points us to time frames in which we can target applicants with marketing campaigns”, “it shows when to focus on various efforts and how to target priorities”, and “it could help us develop strategies to improve application processing efficiencies”. From these contributions, it was clear that most of the participants did not see their perspectives as changed, but all of them concluded that the information was useful to help them improve the work they do. The qualitative contributions indicated that participating in the workshop had not changed participant perspectives while the quantitative data suggested that maybe they had. This is an interesting paradox worthy of investigation but is outside the scope of this research.

Summary – Phase Two

Phase two was designed to answer two research questions aimed at understanding how university enrollment professionals interpret the factors that influence online student enrollment decisions and how that data influences their perspectives, decision making, and strategy development. To answer the research questions, I conducted a workshop coupled with a pre/post-workshop survey to explore the results of the first research phase. The qualitative workshop data was analyzed using inductive coding to extract themes from the workshop/survey responses while quantitative survey data was analyzed using descriptive and inferential statistics.

Prior to receiving the results of the final model, participants ranked each of the 24 factors on a scale from 1-6 then ordered them from 1 to 10 (most to least influential). Overall, the participant’s pre-survey results aligned fairly well with the factors selected in model four, using their own experiences to identify *transfer credits, first-generation* *status, admitted month*, and *Washington residency* as a few of the highest-ranked factors. While they were pretty close to the final model, there was a greater variance in responses on the remaining six factors. Additionally, non-significant factors that ranked high included an *applicant's state, age,* and *military affiliation.*

When presented with the results of the model and the pre-workshop survey, participants' comments fell into two major themes: a desire to understand and a desire for application. As questions, comments, and concerns were expressed to seek clarity, the initial cautious tones of reservation transitioned to optimism, application, and connections to practice. Participants made connections to past strategic planning meetings and connections to future application and decision-making opportunities. Participants made suggestions for the use of the model and expressed interest in taking a deeper dive into the data.

Overall, participants first interpreted the data with caution and curiosity, then optimism and a desire to learn and apply the information to their immediate context. Suggestions to segment the data to focus on learning about specific populations, to change organizational systems to be able to capture and be responsive to live data, and ways to use the information to strategize and prioritize outreach efforts suggest that participants were interested in allowing the information to influence their perspectives, decision making, and strategy development.

Finally, the post-workshop survey was completed and compared to the results of the pre-workshop survey. When ranking the factors on a scale from 1 to 6, participants' responses became much more certain. With few exceptions, the mean scores increased on factors that were included in the final model, and decreased on factors that were excluded, suggesting a slightly higher level of confidence in ranking each factor on the post-survey. While participants responses on the pre-workshop survey hovered near the middle section of the six-point Likert scale, participants seemed to rank factors on the post-workshop survey closer to the extreme ends, indicating an additional sense of certainty in their responses. Additionally, the standard deviations on most of the factors decreased, indicating an increased alignment in participant responses. Using a t test for dependent samples, only the *applied day of the week* and the *admitted day of the week* variables showed significant change with large effect sizes. These variables received the most discussion time during the workshop and seemed to have stuck with the participants when completing the post-workshop survey.

The same sense of certainty came through in the ranking of each factor from 1-10. There were seven non-significant factors included in the top 10 pre-workshop survey responses that were not selected in any of the post-workshop survey top 10. Of the significant factors, 8 of the 10 showed increases in their ranking. Of the nonsignificant factors, 12 of the 14 showed decreases. This demonstrates a higher level of certainty expressed by the participants as the post-workshop survey became more aligned with the model presented during the workshop, further suggesting that participants' perspectives may have been influenced.

Contrary to this finding, participants were asked if their perspectives had changed on which factors influence student enrollment decisions and which do not. Five of the seven responses indicated their perspectives had not changed but provided narrative responses with suggestions on how to use the data. Two indicated their perspectives had changed and provided suggestions on application. All of the participants were also asked if the results of a model that predicts student enrollment decisions can impact enrollment efforts, decision making, and strategy development. All seven participants concluded "yes" and provided a narrative response with application suggestions that mirrored those provided during the workshop.

Summary of Findings

Two research phases were completed with the intent to answer four research questions. Phase one identified the factors that influenced online student enrollment decisions and developed four statistically significant models capable of predicting student enrollment decisions with up to 81% accuracy. Phase two explored how enrollment professionals within the local context interpreted the model and how that information influenced their perspectives, decision making, and strategy development. By and large, most participants' narrative responses on the post-workshop survey indicated their perspectives had not changed but demonstrated clear applications for decision-making and strategy development. While relying on their own experiences, the pre-workshop survey was fairly aligned with the results of the model. However, the post-workshop survey showed further alignment and more certainty in participant responses, suggesting that their perspectives may have changed. Overall, there was consensus that the predictive models of online student enrollment decisions had utility and participants demonstrated a clear desire to understand and apply the model to improve practice within the local context.

CHAPTER 5

DISCUSSION

The purpose of this mixed methods action research study was to understand the factors that influence online undergraduate student enrollment decisions, use those factors to build a model capable of predicting enrollment decisions, and understand how the distribution of this model influences the strategy and decision making of local-level enrollment managers using four-stage theory of organizational change and distributed leadership theory as frameworks for the innovation. In this chapter, I will discuss the findings of each research phase presented in Chapter 4. This chapter also includes a discussion of study limitations as well as recommendations for future practice and research. While this study was conducted within the boundaries of the online campus of a single public research university, the findings will be compared to and discussed with the findings of previous researchers within contexts that differed primarily in student type (traditional vs. nontraditional), institutional setting (online vs. in person), and application of the findings. A discussion of the findings of previous literature as compared to the findings of this study and the theoretical frameworks used to organize the study are woven throughout the remainder of this chapter to illustrate the potential contributions to the available body of literature.

Phase One

As a reminder, the first phase was designed to seek answers to two research questions. The first aimed to understand which factors and to what extent those factors influence the enrollment decisions of admitted, online, undergraduate students. The second was designed to understand how and to what extent those factors predict the enrollment decisions of those students. Using data mining, machine learning models, and informed by previous literature, I narrowed 24 original factors down to 10 independent variables that influenced the enrollment decisions of the student population.

Table 12 illustrates the factors evaluated in this study and compares them to the predictive factors selected by previous studies that shared the same purpose. The student populations and institutional settings are included as important differences that make this research unique. The three previous studies (Basu et al., 2019; Chang, 2006; Luan, 2002) limited their populations exclusively to first-year students or transfer students attending physical campuses, while this study included primarily adult students attending an online campus. Because every study looked at different factors they considered important to their unique populations, only factors evaluated in this study and considered important in one or more of the four studies are included in Table 12. For example, Basu et al (2019) considered *campus visits* as an important predictive factor in their study of first-year students attending a physical campus. However, that factor did not apply to this study, or the data were not available for other studies. Therefore, it was not included in Table 12. Only factors evaluated in this study and considered important in at least one of the other studies are included in the table below.

Table 12

Comparison of Important Factors from Previous Studies

Research	This Study 2023	Basu et al 2019	Chang 2006	Luan 2002
Student Type	Undergraduate Adult Students	Undergraduate First-Year Students	Undergraduate First-Year Students	Undergraduate Transfer Students
Institutional Setting	Online Campus	Physical Campus	Physical Campus	Physical Campus
N	3843	7976	1702	32,000
Factors				
State Residency	✓	✓	✓	
Application Major (Degree Name)	✓	✓	✓	
First-Generation Status	✓	✓		
Student Type	✓		✓	
Total Transfer Credits	✓			✓
Admitted Month	✓			
Admitted Day of the Week	✓			
Degree Type	✓			
Admission Turnaround Time (Days)	✓			
Application Month	✓			
GPA		✓	✓	✓
Gender		✓	✓	✓
Age			✓	✓
Ethnicity			✓	✓
Last School Attended				✓
Zip Code				✓

Significant variables from this study included *Washington residency, admitted month, application major, admitted day of the week, degree type, first-generation status, total transfer credits, number of days it takes to complete the application, application month*, and *student type*. This combination of factors correctly classified the enrollment outcome of 3,846 applicants with 69.2% accuracy. *State residency, application major, first-generation status, student type,* and *total transfer credits* were the only significant factors that aligned with the findings of previous researchers (Basu et al., 2019; Chang, 2006; Luan, 2002), suggesting that these factors may be important across student populations and institutional settings. Meanwhile, the factors that were found to be significant in this study but were not included in any of the previous literature included *admitted month, admitted day of the week, degree type, number of days to complete application*, and *application month*. These factors were not evaluated in previous studies but were included in this study due to practitioner experience. The findings of this study and the variables evaluated can be organized into demographic, geographic, application timing, and academic variables and the findings from each of these groupings are further discussed in this section.

Overall, demographic factors proved to be a surprising finding. An applicant's gender, age, and ethnicity were not significant predictors in this study, although, they were significant factors in most of the previous studies. The only demographic-based factor that was found to be significant was an applicant's *first-generation status* represented by the binary *first-generation* vs. *non-first-generation* categories; this finding was corroborated in the Basu et al. (2019) study. As a result, the data seems to suggest that demographic factors, apart from first-generation students, may differ in importance from studies with different student populations attending physical campuses versus the population of adult online learners included in this study. This finding would need to be further evaluated in similar settings at other adult-serving online institutions to see if it is unique to the local setting or has broader applications.

Demographic factors incorporated into machine learning models create important ethical considerations that researchers must address. Each of the previous researchers suggested that practitioners use predictive models to identify, prioritize, and maximize recruitment efforts. Doing so helps enrollment managers focus their efforts to grow enrollments more efficiently. Consequentially, focusing on populations that demographically align with a particular institution's enrollment history introduces biases that could perpetuate the unintentional exclusion of underserved student populations. If a predictive model indicates that white traditional-aged female students enroll at a much higher rate than Black adult male students, it could be concluded that the university's recruitment efforts could be more effective by targeting the population that enrolls the most efficiently. For this reason, I was surprised that ethnicity, age, and gender were found to be significant and included in previous studies but were not significant factors in this study. Exploring why demographic factors were significant predictors in previous studies but non-significant in this study would require additional research.

Geographic factors that were found to be non-significant included an applicant's *state*, *county*, *city*, *zip code*, *distance from the flagship campus*, and *last school attended*. The only geographic-based factor that was found to be significant was the applicant's residency represented by the binary *Washington resident* vs. *non-resident* variables. Geography did not seem to be an important element for this population even though it was found to be important and significant in previous literature (Basu et al., 2019; Chang, 2006). The data seems to suggest that geographic factors differ from studies whose populations are attending physical campuses versus the population of online learners from this study. This seems to make logical sense as online learners are not limited to geographic locations in the same way on campus learners are.

Factors that measured the influence of application timing had mixed results. Significant factors included the *admitted month, the admitted day of the week, the number of days it takes to complete the application*, and the *application month*. However, other timing-related factors including the *applied day of the week, the number of days applied before the beginning of the semester*, and the *application semester (fall vs. summer)* were all found to be non-significant predictors. There were no particular reasons or information that would suggest the applied day of the week would be important, so the outcome of non-significance was not surprising. However, the *number of days applied before the beginning of the semester* was surprising because this factor seems to be similar to other significant timing-based factors like *application month* or *admitted month*. Application semester (fall vs. summer) was a factor unique to the local context in which applicants who apply to start in summer or fall both are counted as 'fall' applicants. The factor was evaluated to see if there was a difference between their application semester (summer vs fall) and their enrollment outcome. The finding suggests that students can apply to summer or fall semesters without significant changes in their actual enrollment outcome which was surprising and contrary to some of the historical beliefs shared by the research participants in phase two.

Factors related to the timing of a student's application were not included in any of the previous studies. These factors were identified or engineered with existing data because I was familiar with online learners and hypothesized that enrollment yield rates differ based on the timing of a student's application. Before beginning this study, I had a suspicion that timing-based factors could be important, but I did not know if these factors would be significant predictors because they were not included in previous studies. It turns out that many of them are. Students who apply in earlier months enroll at lower rates than students who apply in months closer to the deadline. The faster an application is completed, the more likely they are to enroll. The data also pointed to the *day of the week a student was admitted* as being a significant predictor ($p < 0.001$) with fairly strong odds ratios that ranged from 0.956 to 2.872. This was an interesting factor that was significant in all four models, however, neither I nor the workshop participants in phase two could explain why. As a result, this particular factor requires more exploration to understand how the admitted day of the week data is collected and reported by the office of institutional research to better understand why it was a significant predictor before conclusions can be drawn. Additionally, it could be interesting to explore future cycles that randomly assign the day of the week a student is admitted to determine if this variable remains a significant predictor.

The remaining significant factors found in this study that were also significant in previous literature include an *applicant's major* (degree selection) and *total transfer credits*. The degree selection was a significant predictor in both studies that focused on undergraduate first-year students (Basu et al., 2019 & Chang, 2006), suggesting possible alignment between student types and institutional settings for this predictive factor. *Total* *transfer credit* was a significant predictor in the Luan (2002) study that focused on undergraduate transfer students, suggesting possible alignment between these study's populations and institutional settings. This factor was not evaluated in the other two studies due to the first-year student populations in which transfer credit is not always considered in admissions decisions.

It seems logical that transfer credit has a positive relationship with the enrollment outcome, although only very slightly (Exp(B)) = 1.004). Because this factor has such a low odds-ratio, it should be considered for removal or redesigned for future models. Many of the suggestions from the workshop could be used to re-engineer this factor to make it more useful, including changing this variable to the percent of transferable credits or the proportion of transfer credits that applied towards degree completion. Since adult learners often return to college after an extended absence, the time to degree completion is often at the top of their list. My experience has found that online adult learners will often select a school that offers them the most valuable transfer credit and the shortest time to degree completion, so engineering these variables to test that hypothesis might provide value and strengthen the model in future cycles.

The remaining non-significant factors from this study included an applicant's *GPA* and their *last school attended*. Discovering *GPA* was not a significant factor in this study was surprising because it is considered a standard tool used to categorize and model future enrollment trends by all of the previous researchers and the office of intuitional research within this local context. At the university where this study took place, the office of institutional research currently produces new-student enrollment forecasts using GPA as the primary factor because they have found this factor to be the most accurate predictor of enrollment outcomes. By categorizing and calculating the yield rates for previous years of applicants with similar GPAs (examples: 3.0-3.1 GPA, 3.11-3.2 GPA, 3.21-3.3 GPA, etc.), they project the enrollment outcomes of an existing population of admitted students based on the same GPA categories. For example, if applicants with GPAs between 3.0 and 3.1 enroll at 54% over the last three years, that logic is then forecasted forward to the current pool of applicants within that GPA category.

Similar to the previous studies, the university in this study primarily serves first-year student populations attending one of five physical campuses. As the only online campus within the university system, I found that GPA was not a significant predictor of student enrollment. Therefore, this particular finding can be shared with the office of institutional research to improve system-wide enrollment forecasting. GPA was considered important to both first year and transfer populations attending physical campuses in the previous literature, but it was not important for the online adult students within this study. This is an important finding that could be further evaluated to see if it applies to just this context or more broadly to other online campuses or institutions.

Finally, other studies included important variables that were not easily available within the local context during the study period. There are significant factors from previous studies that would be interesting to evaluate within the local context of this study or the broader context of adult serving online institutions. These factors include a student's legacy status, application source (inquiry vs stealth app[1]), event participation (example: attended pre-enrollment orientation), financial aid status, communication frequency and modality, employment status, number of previous schools attended,

[1] An applicant is considered a stealth app when their first contact with the university was their application.

scholarship amount, time away from school, number of interactions with an admissions counselor, and the expected hours per week spent on extra-curricular obligations. These factors are examples of variables I was unable to collect but would serve as interesting variables to evaluate in future cycles of research.

Phase Two

Phase two was designed to accompany the first phase and extend the scope of previous research to seek answers to two additional research questions. Guided by the four-stage theory of organizational change framework and distributed leadership theory, the first research question aimed to understand how university enrollment professionals interpret the factors that influence online student enrollment decisions. The second question was designed to understand how that knowledge affects enrollment professionals’ perspectives, decision making, and strategy development. Research participants engaged in a workshop accompanied by a pre/post workshop survey. This phase was guided and informed by distributed leadership theory, was designed using the four-stage theory of organizational change and produced both quantitative survey data and qualitative workshop data.

Previous researchers with similar studies ended their research with suggestions on how practitioners could consider using machine learning to predict student enrollment decisions. The outcome of their research was to illustrate that predictive modeling was valid and had a place in higher education enrollment management. However, they all stopped short of connecting their research findings back to the local settings and discussing the findings with practitioners to evaluate their actual interpretations and use of the information. The purpose of phase two was to extend existing research to have practitioners discuss and apply the findings of the model within the local context.

As discussed in chapter four, two major themes were derived from the qualitative data. Participants shared a desire to understand/learn more and a desire to apply the knowledge to inform practice, strategy, and decision making. The workshop was intentionally designed using distributed leadership theory, which posits that the work of leadership takes place across many people in complex organizational settings (Bolden, 2011). The goal of the workshop was to create a space for which leadership decision making could occur as a social process with cross-organizational stakeholders. Their contributions and the resulting themes extracted from the workshop followed the four-stage theory of organizational change. Participants engaged in conversation to collaboratively evaluate the utility of the model and its potential for use within each of their professional practices. Participants moved through the first two stages of becoming familiar with the problem and possible solutions (stage 1) then discussing adoption (stage 2). These stages aligned with the qualitative codes and extracted theme, which indicated that participants had a desire to understand and learn more about the data and how it applies to their work. As participants became familiar with the data and the model, the qualitative data codes continued to follow the last two stages of this theory. Participants began providing suggestions on how to modify organizational structures to use the new information (stage 3) and suggestions on how to incorporate the information in the ongoing activities through strategic planning (stage 4). Their contributions and the qualitative codes resulted in the extracted theme that participants had a desire to apply the information to inform organizational practice, strategy, and decision making.

While the detailed discussion of these themes and their connection to theory was provided in chapter four, this section explores the implications of discovering these themes. Participants repeatedly expressed a desire to understand each of the variables and explore them much deeper. The presentation of the binary outcomes, the *X* variable was significant while the *Y* variable was not significant, did not seem to satisfy their curiosity. For example, one participant expressed surprise that ethnicity was not a predictive factor. They followed up their surprise with a series of questions wanting to know which ethnicities were included, what were each of their individual yield rates, and what were the sizes of each population? This line of inquiry was true for most of the factors. They did not just want to know that *admitted month* was significant, they wanted to know why, which months were better than others, and how that informs the work we do.

Due to the limited time of the scheduled workshop, we were unable to explore each factor in the level of detail the participants were interested in. The four-stage theory of organizational changed is used to describe how change evolves over time. However, this framework was used within this study to facilitate an innovation that followed those stages, which resulted in an accelerated version in which each stage occurred within a short period of time. However, the desire to explore some of the more complex factors led to several ‘offline’ conversations and data sharing among the participants in the weeks that followed the workshop. Discovering a factor was a significant predictor was interesting, but it was not enough information to satisfy their desire to learn. Participants wanted to learn more, understand the information, segment it to particular populations, and use it to inform their professional practice. These outcomes were facilitated through conceptional perspective of distributed leadership theory and aligned with the four-stage theory of organizational change. Many of the participants' questions and suggestions could easily lead to future cycles of action research and are explored in the recommendations for future research section of this chapter.

The quantitative pre/post survey data presented in chapter four illustrated that participants' perspectives likely changed as a result of participating in the workshop and learning about the predictive factors of online student enrollment decisions. However, their narrative survey responses seemed to tell a different story. Likert scale scores and ranking orders gravitated more toward the extreme ends on the post-workshop survey, suggesting that participants became more certain in their responses. Twelve of the fourteen non-significant variables showed decreases in ranking scores on the post survey while eight of the ten significant variables showed increased scores. This illustrates that the participants' perspectives became more aligned with the predictive model after participating in the workshop.

As described in chapter four, five of the seven narrative survey responses indicated their perspective had not changed when asked *has your perspective changed on which factors influence student enrollment decisions and which do not?* Participants expressed that they felt like they had more questions they wanted answered. While the survey showed some signs of changed perspectives, I believe the way participants viewed their own perspectives did not seem to change because they were more interested in understanding the information in greater detail. This finding aligned with the themes identified from the qualitative data and led to many additional 'offline' discussions after the workshop. Additionally, the accelerated timeline of progression through the four-stage theory of organizational change could have contributed to this paradox as well.

Participants seemed to focus on the areas they wanted to learn more about, rather than the parts they learned and clearly understood. The relatively short timeframe for discussion, coupled with a focus on the topics they did not understand, provides insights into why participants reported their perspectives had not changed.

Although the survey data suggested one thing while the narrative survey responses suggested another, I believe that participating in a 90-minute workshop was not enough time for participants to view their own perspectives as different. Instead, they left wanting more information and future conversations to learn more about our student population to use that knowledge to change and inform practice. The workshop really served as an introduction and a catalyst to future discussions through distributed leadership theory, long-term change, and future cycles of inquiry though a prolonged application of the four-stage theory of organizational change.

Limitations

The purpose of action research is to learn more about one's own practice to better understand and improve the quality or effectiveness of the work we do within a specific setting (McMillan, 2004; Mertler, 2020; Schmuck, 1997). Because the research is focused within a specific setting, it is not intended to have broader applications. Additional research would be required to draw generalizable conclusions for broader populations. The references to differences between each student population and institutional setting from previous studies are described for comparison purposes only and not intended to convey a broader application. These comparisons and differences, though, function as interesting topics for future researchers to explore if consistency can be achieved across various student populations and institutional settings.

Additionally, I played dual roles as both the researcher and a member of the enrollment management team within the study. Existing relationships and the desire to maintain face could have played a role in how the participants engaged in the study. The qualitative data in phase two was also evaluated and analyzed by the researcher, a member of the enrollment management team. While every effort was made to address and remove positive bias, it is certainly possible that the dual role introduced unintentional bias. That being said, every attempt was made to report all findings in an honest and transparent way. It is possible that an unrelated third-party researcher without existing relationships with participants could have executed and analyzed the data from phase two and discovered different results.

Previous researchers included variables that were not easily available or were simply not captured within the local context during the study period (Basu et al., 2019; Chang, 2006; Luan, 2002). Some of these variables include an applicant's legacy status, financial aid data, scholarship amount, and event participation. Other variables that were unavailable included communication frequency and modality, employment status, attitudes towards higher ed, number of previous schools attended, time away from school, number of interactions with an admissions counselor, and the expected hours per week spent on extra-curricular obligations. The inability to access these variables presented a limitation to this study and an opportunity for future cycles after adjustments to organizational processes are made.

Chapter four presented a paradox that showed that narrative responses to the survey indicated participants mostly believed their perspectives had not changed as a result of participating in the workshop while the changes in pre/post survey data suggested that a change may have occurred. While this paradox is an interesting finding, the reasons for its existence are outside the scope of this research and would require additional exploration to understand exactly why it occurred. I certainly will not claim to know exactly why this paradox occurred; however, I believe there are several possible reasons for this that have to do with the study's time constraints, existing relationships among the participants, and long-held positions on specific topics related to the distributed leadership environment within the local context. Due to the limited time available to conduct the study, the stages of change were measured within a relatively short window of observation. This short time frame could have contributed to the paradox between the qualitative reporting that participant perspective had not changed and the quantitative evidence that they did change. It would be interesting to measure the long-term progression and application of the theories that informed stage two.

Distributed leadership theory was used to describe the context and social environment within the study's setting. This theory posits that the work of leadership takes place in deliberative settings across multiple people especially in educational organizations instead of focusing on the characteristics of individual leaders. Leadership in this environment is seen as a social process that involves everyone including leaders and followers. As a result, existing relationships, different levels of authority, and long held beliefs are all possibilities of factors that may have played a role in why this paradox existed. The influence of possible resistance to change are magnified in situations where leadership is shared because individual voices are amplified and carry power that may not exist in traditional leadership settings where decisions are made by individual leaders. This is especially true when topics and findings challenge long-held beliefs which can be difficult to accept and understand for individuals and groups alike. Ultimately, there are several explanations that could help understand why the paradox existed between the quantitative and qualitative results of phase two. More exploration would be required to determine exactly why this finding existed.

The GPA and address variables both contained missing data that required different forms of imputation to remain included in the final data set. While the imputation methods I applied followed the choices of previous literature (Basu et al., 2019), imputation introduces bias and impacts variance (Kang H, 2013; Khan S & Hogue, A, 2020). Imputation may lead to inconsistent bias and an underestimation of errors. Using imputation methods allowed for the continued use of the GPA variable (n=424) and address variable (n=24), however, provides a noteworthy limitation to these specific variables and to this study.

Additionally, categorical variables that have a small number of values present the possibility of over interpretation of the sign of estimates. While the distribution of each categorical variable is disclosed in Table 4, this issue of unbalanced categories creates a potential limitation to the interpretation of the model's effectiveness.

Finally, the sample size from phase two only included seven enrollment professionals with various backgrounds and levels of seniority. Existing relationships and reporting structures (e.g., workshop participants included management and direct reports) could have influenced how a participant engaged in the workshop. Because of the limited sample size, changes in pre/post survey responses should be interpreted with caution. Additionally, while all seven participants completed the pre-workshop survey ranking question (order from 1 to 10), only five completed the ranking question on the post-workshop survey. This small gap presents a limitation to the utility of that data set.

The research conclusions provided as a result of completing this study were reached using data and experience specific to the local context. More research is required to support a broader application of the findings to larger populations, and therefore, future researchers should only apply these findings to their own settings with great caution. I believe it is safe to conclude that predictive models have a place in enrollment management practices in higher education, but the unique settings of each institution should be considered when interpreting the results of this study.

Recommendations for Future Research

This research study presented nearly countless ways to conduct future cycles of action research. Participants expressed a desire to build individual models for each student type in order to further segment and inform recruitment strategies for different populations. Participants made suggestions to optimize the model by evaluating additional variables such as financial data, engineering variations of transfer credit data, and broadening the population to include other start terms. Participants suggested we consider building a similar model to not only predict enrollment decisions, but also begin to predict student retention from the admissions stage of a student journey. Each of these could easily contribute to future cycles of action research and improve practices within the local context.

The broadest suggestion I can make is to encourage future researchers to explore adult student enrollment decisions at other online institutions. There are multiple examples of studies that focus on the enrollment decisions of traditional-aged students attending physical campuses and those studies have mostly aligned on similar factors and machine learning methods. However, I was unable to find a study that focused on adult learners in online university settings. This gap in the available literature provides an opportunity to learn more about the enrollment decisions of adult learners, support enrollment professionals, and ultimately improve the services we provide to future generations of students. More research in this area at other universities could provide a foundation for broader application and suggestions for improved practice that are publicly available.

This study used application data to predict student enrollment decisions, then shared the results with a group of enrollment managers. Another avenue I believe would be worthwhile to explore is collecting qualitative data directly from the students. Designing interviews or surveys that prompted students to answer *why did you choose to attend this university* or *why did you choose not to enroll in classes* would provide an additional layer of data that could be interesting to explore. Many institutions have a mechanism for capturing information when a student leaves, like issuing an exit survey, but I am not familiar with many institutions that ask their students, why did you pick us? This extra cycle of inquiry could provide qualitative data to help further understand the enrollment behaviors of any student population.

Online adult students have motivations for returning to higher education that are different than those of traditional-aged college students attending physical campuses. Experience has shown that prospective adult learners have three basic questions that nearly all ask during informational interviews: 1) how long will it take, 2) how much will it cost, and 3) will it help me achieve my goals? I believe there is an opportunity to incorporate finance data into future iterations of predictive models. This could include financial aid information, socioeconomic status, cost of attendance information, and expected family contribution, among other financial data points. Unfortunately, financial data was considered sensitive data that I was unable to access during the study period creating both a limitation to this study and an opportunity for future cycles.

Additionally, engineering transfer credit data in a way that illustrates the time to degree completion seems like an opportunity. Communicating these findings back to prospective students could also provide benefits. I suspect that adult students who are closer to finishing will enroll at a higher rate than those that have longer completion timelines, but this hypothesis requires additional testing which the enrollment professionals in the local context have committed to learn more about. Engineering features and building organizational structures to capture this data would be an interesting opportunity. Finally, creating features that capture and categorize the motivations of the prospective adult students presented in chapter three could provide value. For example, is a student returning for economic or career reasons, personal fulfillment, role modeling for children, or other reasons? If we can identify, categorize, and collect motivations, they could also be used for modeling enrollment behaviors.

The enrollment professionals in the local context agreed that developing a way to capture applicant narratives why they chose to or not to enroll would increase our understanding of the student enrollment decision process and inform future interventions. This project was adopted shortly after the research period with plans to survey both applicant populations and build interventions based on survey responses. This is an example of a future cycle of action research to further evaluate this topic and increase practitioner understanding of our local context.

The data was collected from internal data sources by local systems within the study context. There is also an opportunity to utilize external data sources to expand the scope of this research and the findings. For example, the National Student Clearinghouse provides their membership with data on where students ultimately enrolled. Within the local context, this is referred to as ‘lost market data’ and essentially tells us if an applicant enrolled at a different college or university. This external data source, among other publicly available data sources through the national center for educational statistics, could provide interesting factors for both local level application and possible broader extrapolation.

More research into the enrollment decisions of online adult learners at other institutions or with national level external data could lead to broader applications and publicly available findings. New sources of data including asking the students to respond to why they chose or did not choose to enroll could provide new perspectives into their enrollment decisions in ways we have yet to consider. And building organizational systems to capture, incorporate, and respond to new features we believe could be important to our population of students could lead to more accurate and optimized predictive models. Each of these suggestions aims to better understand the students we work with and ultimately improve the way institutions of higher education serve their student populations.

Conclusion

Student enrollment in institutions of higher education has declined nationwide every year since 2010 (U.S. Department of Education, 2021). Since 2019, the number of adult learners in the United States with some college credit but no degree and are no longer enrolled has increased from 36 million to 40.4 million (Causey et al., 2023). Scholars and consumers are increasingly questioning the value of a college degree as access to low-cost educational alternatives has become commonplace (Blanco Ramirez & Berger, 2014; Jones, 2013; Tomlinson, 2018). To address enrollment challenges and guide institutions to long-term sustainability, university administrators are tasked with improving their understanding of the students they serve and their place in the knowledge economy. While educational data mining and machine learning will not solve these challenges alone, these tools provide a way for enrollment leaders to understand their student populations and shape programs, resources, and activities that optimize efficiency and improve student services.

This study builds upon a body of literature that supports the use of machine learning to organize an infinitely growing supply of data into insightful and actionable knowledge that helps combat the challenges faced in higher education. These methods go well beyond the scope of this study. Machine learning can be applied to identify and intervene with at-risk students to reduce student attrition. Machine learning can be used to identify, categorize, and target outreach to alumni and potential donors with the right messages and most effective donation amounts to optimize campaign results. This study illustrates that machine learning can be used to predict the enrollment decisions of online adult students and that enrollment managers have a genuine desire to understand and use the data to combat the many challenges they are tasked to overcome. Machine learning has a place in higher education and the possibilities for its application are only limited to the imaginations of future researchers.

REFERENCES

Al-Twijri, M. & Noaman, A. (2015). New data mining model adopted for higher institutions. *Procedia Computer Science*, *65*, 836 – 844. https://doi.org/10.1016/j.procs.2015.09.037

Anderson, G. (2002). Reflecting on research for doctoral students in education. *Educational Researcher, 31*(7), 22 – 25. https://doi.org/10.3102/0013189X031007022

Aslanian, C. B. (2001). *Adult students today*. The College Board.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, *1*(1), 3 – 17. https://doi.org/10.5281/zenodo.3554657

Basu, K., Basu, T., Buckmire, R., & Lal, N. (2019). Predictive models of student college commitment decisions using machine learning. *Data (Basel), 4*(2), 65. https://doi.org/10.3390/data4020065

Bennett, N., Wise, C., Woods, P., & Harvey, J. (2003). *Distributed leadership: A review of literature*. National College for School Leadership.

Blanco Ramirez, G., & Berger, J. (2014). Rankings, accreditation and the international quest for quality: Organizing an approach to value in higher education. *Quality Assurance in Higher Education*, *22*(1), 88 – 104.

Bolden, R. (2011). Distributed leadership in organizations: A review of theory and research. *International Journal of Management Reviews, 13*(3), 251 – 269. https://doi.org/10.1111/j.1468-2370.2011.00306.x

Bruggink, T. H., & Gambhir, V. (1996). Statistical models for college admission and enrollment: a case study for a selective liberal arts college. *Research in Higher Education, 37*(2), 221 – 240.

Causey, J., Gardner, A., Pevitz, A., Ryu, M., & Shapiro, D. (2023). *Some college no credential student outcomes, annual progress report – Academic year 2021/2022.* National Student Clearinghouse Research Center.

Chang, L. (2006). Applying data mining to predict college admissions yield: A case study. *New Directions for Institutional Research,* (131), 53–68. https://doi.org/10.1002/ir.187

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0.* Step-by-step data mining guide.

Choy, S. (2002). *Findings from the condition of education 2002: Nontraditional undergraduates.* U.S. Department of Education, National Center for Education Statistics. https://files.eric.ed.gov/fulltext/ED546117.pdf

Corley, M. A. (2022). Adult learning theories. *Teaching excellence in adult literacy*. TEAL Center Fact Sheet No. 11. https://lincs.ed.gov/state-resources/federal-initiatives/teal/guide/adultlearning

Creswell, J. (2005). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research.* Merrill/Prentice Hall.

Creswell, J. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). Sage.

Creswell, J., & Guetterman, T. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.

Creswell, J., & Plano Clark, V. (2011). *Designing and conducting mixed methods research.* Sage.

Delen, D. (2011). Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory & Practice, 13*(1), 17 – 35. https://doi.org/10.2190/CS.13.1.b

Dumais, S., Rizzuto. T., Cleary, J., & Dowden, L. (2013) Stressors and supports for adult online learners: Comparing first- and continuing-generation college students. *American Journal of Distance Education, 27*(2), 100-110. https://doi.org/10.1080/08923647.2013.783265

Duzhin, F., & Gustafsson, A. (2018). Machine learning-based app for self-evaluation of teacher-specific instructional style and tools. *Education Sciences, 8*(1), 7. doi:10.3390/educsci8010007

Fischer, K., & Ellis, L. (2021, January 27). The heavy cost of an empty campus: Decades of disinvestment left public research universities overexposed to Covid-19. *The Chronicle of Higher Education, 67*(11), 5. https://go-gale-com.ezproxy1.lib.asu.edu/ps/i.do?p=OVIC&u=asuniv&id=GALE%7CA652867395&v=2.1&it=r&aty=ip

Garrett, R. (2021). *The new higher education: College alternatives are good news for students and colleges.* Eduventures Summit.

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). *Health behavior and health education: Theory, research, & practice (4th ed.).* Jossey-Bass.

Gronn, P. (2000). Distributed properties: A new architecture for leadership. *Educational Management & Administration, 28*(3), 317 – 338. https://doi.org/10.1177/0263211X000283006

Gronn, P. (2002). Distributed leadership as a unit of analysis. *The Leadership Quarterly, 13*(4), 423 – 451. https://doi.org/10.1016/S1048-9843(02)00120-0.

Grus, J. (2015). *Data science from scratch first principles with python*. O'Reilly. https://doi.org/10.1016/S1048-9843(02)00120-0

Guabassi, I., Bousalem, Z., Marah, R., & Qazdar, A. (2021). A recommender system for predicting students' admission to a graduate program using machine learning algorithms. *International Journal of Online and Biomedical Engineering, 17*(2), 135 – 147. https://doi.org/10.3991/ijoe.v17i02.20049

Herr, K., & Anderson, G. (2005). *The action research dissertation: A guide for students and faculty.* Sage.

International Educational Data Mining Society. (2022). *Educational data mining*. https://educationaldatamining.org/

Ivankova, N. (2015). *Mixed methods applications in action research: From methods to community action.* Sage.

Jones, G. (2013). Rates of exchange: Neoliberalism and the value of higher education. *International Studies in the Sociology of Education, 23*(3), 273 – 280.

Kang, H. (2013). The prevention and handling of missing data. *Korean Journal of Anesthesiology, 64*(5), 402 – 406. https://doi.org/10.4097/kjae.2013.64.5.402

Khan, S., & Hoque, A. (2020). SICE: An improved missing data imputation technique. *Journal of Big Data, 7*(37). https://doi.org/10.1186/s40537-020-00313-w

Kimmel, S. B., & McNeese, M. (2006). The influence of gender, ethnicity, and age on adult motivation and barriers to education. *International Journal of Diversity in Organizations, Communities & Nations, 5*(1), 33 – 41.

Knowles, M. (1980). *The modern practice of adult education: Andragogy versus pedagogy*. Cambridge Adult Education.

Knox. (2019). How state disinvestment has hurt public universities. *The Chronicle of Higher Education, 65*(36), A16.

Livieris, K., Tampakas, V., & Pintelas, P. (2018). Improving the evaluation process of students' performance utilizing a decision support software. *Neural Computing & Applications*, *31*(6), 1683–1694. https://doi.org/10.1007/s00521-018-3756-y

Luan, J. (2002). Data mining and its applications in higher education. *New Directions for Institutional Research, 2002*(113), 17 – 36. https://doi.org/10.1002/ir.35

Magda, A. J., Capranos, D., & Aslanian, C. B., (2020). *Online college students 2020: Comprehensive data on demands and preferences*. Wiley Education Services. Retrieved from https://universityservices.wiley.com/wp-content/uploads/2020/10/Report-2020-Online-College-Students.pdf

Mallikharjuna, K., Saikirishna, G., & Supriya, K. (2023). Data preprocessing techniques: Emergence and selection towards machine learning models – a practical review using HPA dataset. *Multimedia Tools and Applications, 82*(24), 37177 – 37196. https://doi.org/10.1007/s11042-023-15087-5

Mattern, K. & Wyatt, J. (2009). Student choice of college: How far do students go for an education? *Journal of College Admission, 203,* 18-29.

McMillan, J. (2004). *Educational research: Fundamentals for the consumer.* Allyn & Bacon.

Mertler, C. (2020). *Action research: Improving schools and empowering educators.* Sage.

Millar, R.B. (2011). *Maximum likelihood estimation and inference: With examples in R, SAS and ADMB* (1st ed.). Wiley.

Mills, G. (2011). *Action research: A guide for the teacher researcher.* Pearson.

Nascimento, R., Neves, R., Almeida Neto, M., & Araújo Fagundes, R. (2018). Educational data mining: An application of regressors in predicting school dropout. *Machine Learning and Data Mining in Pattern Recognition, 10935,* 246 – 257. https://doi.org/10.1007/978-3-319-96133-0_19

National Center for Educational Statistics. (n.d.). *Definitions and data: Who is nontraditional?*. U.S. Department of Education. Retrieved from https://nces.ed.gov/pubs/web/97578e.asp

National Center for Education Statistics. (2021). Undergraduate enrollment. *Condition of Education*. U.S. Department of Education, Institute of Education Sciences. Retrieved from https://nces.ed.gov/programs/coe/indicator/cha.

Osgood-Treston, B. (2001). Program completion barriers faced by adult learners in higher education. *Academic Exchange Quarterly, 5*(2), 120.

Parry, K. W., & Bryman, A. (2006). Leadership in organizations. In S. R. Clegg, C. Hardy, T. Lawrence, & W. R. Nord (eds.), *The SAGE handbook of organizational studies* (pp. 447 – 468). Sage.

Reips, U. (2010). Design and formatting in internet-based research. In S. Gosling & J. Johnson (eds.), *Advanced methods for conducting online behavioral research* (pp. 29 – 43). American Psychological Association. https://doi.org/10.1037/12076-003

Rogers, E. (2003). *Diffusion of Innovations* (5th ed.). Simon & Schuster.

Romero, C. & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135–146. https://doi.org/10.1016/j.eswa.2006.04.005

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (2010). *The handbook of educational data mining.* CRC Press.

Ross-Gordon. (2011). Research on adult learners: Supporting the needs of a student population that is no longer nontraditional. *Peer Review: Emerging Trends and Key Debates in Undergraduate Education, 13*(1), 26 – 29.

Roueche, J., & Roueche, S. (2000). High stakes, high performance: Making remediation work. *American Association of Community Colleges.*

Schmuck, R. (1997). *Practical action research for change.* Skylight Professional Development.

Schneier, B. (2015). *Data and Goliath: The hidden battles to collect your data and control your world*. W.W. Norton & Company.

Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019). Student performance prediction and classification using machine learning algorithms. *Proceedings of the 2019 8th International Conference on Educational and Information Technology, 148151,* 7 – 11. https://doi.org/10.1145/3318396.3318419

Shapiro, D., Ryu, M., Huie, F., & Liu, Q. (2019). *Some college, no degree: A 2019 snapshot for the nation and 50 states*. Signature Report No. 17, National Student Clearinghouse Research Center.

Smart, J. (2005). Attributes of exemplary research manuscripts employing quantitative analyses. *Research in Higher Education, 46*(4), 461-477.

Tampakas, V., Livieris, I., Pintelas, E, Karacapilidis, N, Pintelas, P. (2018). *Prediction of students' graduation time using a two-level classification algorithm.* In Proceedings of the 1st International Conference on Technology and Innovation in Learning, Teaching and Education.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences.* Sage.

Tomlinson, M. (2018). Conceptions of the value of higher education in a measured market. *Higher Education, 75*(4), 711 – 727. https://doi.org/10.1007/s10734-017-0165-6

U.S. Department of Education, National Center for Education Statistics, (2021). *Fall enrollment in colleges and universities.* Integrated Postsecondary Education Data System (IPEDS). https://nces.ed.gov/ipeds/TrendGenerator/app/build-table/10/101?rid=2&cid=102

U.S. Department of Education, National Center for Education Statistics (2020). *Fall Enrollment component final data (2003 - 2019)*. Integrated Postsecondary Education Data System (IPEDS). https://nces.ed.gov/ipeds/TrendGenerator/app/trend-table/2/8?trending=column&rid=83

Vartto. (2019). Democratization of organizational change process through deliberation. *Leadership & Organization Development Journal*, *40*(7), 791 – 802. https://doi.org/10.1108/LODJ-02-2019-0066

Washington State University (2022). *Quick facts.* https://wsu.edu/about/facts/

Waters, A., & Miikkulainen, R. (2014). Grade: Machine-learning support for graduate admissions. *The AI Magazine*, *35*(1), 64–75. https://doi.org/10.1609/aimag.v35i1.2504

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge University Press.

Yağcı. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments, 9*(1), 1 – 19. https://doi.org/10.1186/s40561-022-00192-z

Yates, H. & Chamberlain, C. (2017). *Machine learning and higher education*. EDUCAUSE. https://er.educause.edu/articles/2017/12/machine-learning-and-higher-education

Yukselturk, E., Ozekes, S., & Türel, Y. (2014) Predicting dropout student: An application of data mining methods in an online education program. *European Journal of Open, Distance, and E-Learining*,*17*(1), 118–133. DOI:10.2478/eurodl-2014-0008

APPENDIX A

PRE/POST WORKSHOP SURVEY

Pre-Workshop Survey

To protect your confidentiality, please create a unique identifier known only to you. To create this unique code, please record the first three letters of your mother's first name and the last four digits of your phone number. Thus, for example, if your mother's name was Sarah and your phone number was (602) 543-6789, your code would be Sar 6789. The unique identifier will allow us to match your post-intervention survey responses and your retrospective, pre-intervention responses when we analyze the data.

My unique identifier is: _____ ________ (e.g., Sar 6789, see paragraph above)

Section 1 – Demographic Questions:

1. How long have you worked professionally in higher education?
    a. 0-5 years
    b. 6-10 years
    c. 11-15 years
    d. 16-20 years
    e. 20+ years
2. How much of your career in higher education has been in enrollment management?
    a. 0-5 years
    b. 6-10 years
    c. 11-15 years
    d. 16-20 years
    e. 20+ years
3. How much of your career in higher education involved working directly with students on a daily basis?
    a. 0-5 years
    b. 6-10 years
    c. 11-15 years
    d. 16-20 years
    e. 20+ years
4. How much of your career in higher education involved working indirectly with students (less than a daily basis)?
    a. 0-5 years
    b. 6-10 years
    c. 11-15 years
    d. 16-20 years
    e. 20+ years
5. How many years has the majority of your work been focused on serving online adult learners?
    a. 0-5 years
    b. 6-10 years
    c. 11-15 years
    d. 16-20 years
    e. 20+ years
6. Are you?

   a. Male
   b. Female
   c. Prefer not to answer
7. What rase or ethnicity best describes you?
   a. American Indian or Alaskan Native
   b. Asian / Pacific Islander
   c. Black or African American
   d. Hispanic
   e. White / Caucasian
   f. Multiple Ethnicity / Other
   g. Prefer not to answer

Section 2 – Select the Level of Influence for Each Factor

Directions: For the following sections, please indicate the level you believe each factor may influence the enrollment outcome of an admitted student on a six-point scale.
For example, if you believe a student's shoe size has a strong influence on if they will enroll, you would select "6 – Strong Influence". If you believe that a student's hair color has no influence on their enrollment outcome, you would select "1 – No Influence". If you believe there may be some level of influence, select the best answer from levels 2-5 that best reflects your perspective.

Demographic Information.
I believe that an admitted student's ___*(factor)*___ has ___*(level of influence)*___ on their likelihood to become an enrolled student.

	1 – No Influence	2	3	4	5	6 – Strong Influence	*Prefer not to answer*
Age	○	○	○	○	○	○	○
Ethnicity	○	○	○	○	○	○	○
First-generation Status	○	○	○	○	○	○	○
Gender	○	○	○	○	○	○	○
Military Affiliation	○	○	○	○	○	○	○

Student Application Information.
I believe that an admitted student's ___*(factor)*___ has ___*(level of influence)*___ on their likelihood to become an enrolled student.

	1 – No Influence	2	3	4	5	6 – Strong Influence	*Prefer not to answer*
Application Major (Degree Name)	○	○	○	○	○	○	○
Applicant's Cumulative GPA	○	○	○	○	○	○	○

Degree Type (first undergraduate degree, post-baccalaureate, or non-degree seeking)	○	○	○	○	○	○	○
Student Type (transfer, first year, former student returning, or non-degree seeking)	○	○	○	○	○	○	○
Total Number of Transfer Credits	○	○	○	○	○	○	○
Last School Attended	○	○	○	○	○	○	○

Geographic Information.
I believe that an admitted student's ___*(factor)*___ has ___*(level of influence)*___ on their likelihood to become an enrolled student.

	1 – No Influence	2	3	4	5	6 – Strong Influence	*Prefer not to answer*
City	○	○	○	○	○	○	○
County	○	○	○	○	○	○	○
State	○	○	○	○	○	○	○
Zip Code	○	○	○	○	○	○	○
Washington Residency	○	○	○	○	○	○	○
Distance from Pullman Campus (miles	○	○	○	○	○	○	○

Application Timing Information.
I believe that the ___*(factor)*___ has ___*(level of influence)*___ on an admitted student's likelihood to become an enrolled student.

	1 – No Influence	2	3	4	5	6 – Strong Influence	*Prefer not to answer*
Application Semester (Summer vs Fall)	○	○	○	○	○	○	○
Day of the Week Student Submitted Admissions Application (Monday, Tuesday…etc.)	○	○	○	○	○	○	○
Day of the Week Student was Admitted (Monday, Tuesday…etc.)	○	○	○	○	○	○	○
Month Student Submitted Application for Admission	○	○	○	○	○	○	○

Month Student was Admitted	○	○	○	○	○	○	○
Number of Days between 'Application Submitted' and 'Admitted'	○	○	○	○	○	○	○
Number of Days between 'Application Submitted' and 'First Day of Classes'	○	○	○	○	○	○	○

Section 3 – Ranking Factors

Directions: From the list below, select the top ten (10) factors that you believe have the strongest influence on if an admitted applicant will enroll or not. Then, order them from 1 (most influence) to 10 (least influence).

1. {List All Factors}.
2. …

Section 3 – Almost Finished

1. Are there factors that were not included in this study you believe would influence the enrollment outcome of an admitted student?
   a. Yes
      i. Which factors do you suggest (open textbox response)
   b. No

Post-Workshop Survey

The post post-intervention survey was identical to the pre-intervention survey, except:

1. Post-Intervention Survey:
   a. Omitted demographic questions (section 1)
   b. Added the following open-ended questions at the end of the survey:

Please answer the questions below with a narrative response.

1. Which factors supported your beliefs (i.e. which where you correct about) and why did you select them?
2. Were there any factors that surprised you (i.e. a factor that was not predictive when you though it would be, or vice verse)?
   a. Yes
      i. Which factors were you surprised by?
   b. No
3. Has your perspective changed on which factors influence student enrollment decisions and which do not? How so, or why not?

4. Can the results of a model that predicts student enrollment decisions impact enrollment efforts, decision making, and strategy development? How so, or why not?
5. Based on your experience, which factors (if any) were not considered that you believe should be included and would improve the utility of the predictive model?
6. Is there anything else you would like to add?
   a. Yes
      i. What would you like to share?
   b. No

APPENDIX B

ONE-PAGE MODEL & PRE-SURVEY SUMMARY

Predicting Online Undergraduate Student Enrollment Decisions: A Workshop

Research Summary

The purpose of this study is to understand what factors (variables) influence online student enrollment outcomes at Washington State University Global Campus (WSUGC) and how enrollment professionals can use this information to improve services for online students. The study evaluated 24 unique factors from 3843 applicants who were admitted to WSUGC for summer or fall semesters between 2020 and 2022.

Table 1 – Predictive Factors — **Survey Results**

	Student Factors	Mean	SD	% of Responses Selected in Top 10
1	Student Type (transfer, first year, returning, non-degree)***	4.14	1.25	43%
2	WA Residency***	4.86	0.99	71%
3	Admitted Month***	4.71	0.88	86%
4	Application Major (Degree Name)***	3.14	1.12	14%
5	Admitted Day of the Week***	2.29	1.16	14%
6	Degree Type (first undergraduate, post-bacc, non-degree)***	4.29	1.28	43%
7	Application Month***	4.71	0.88	43%
8	First Generation Status*	4.5	0.9	71%
9	Number of days between 'App Submitted' to 'Admitted'*	5	0.76	57%
10	Total Number of Transfer Credits***	5	0.76	86%

Table 2 – Non-Predictive Factors — ***Survey Results***

	Student Factors	Mean	SD	% of Responses Selected in Top 10
1	Number of days between 'App Submitted' and the 'First Day of Classes'	4.57	0.73	100%
2	Applicant's Cumulative GPA (from previous schools)	4.14	1.12	57%
3	State	4.86	0.83	43%
4	Age	4.57	1.05	43%
5	Military Affiliation	4.43	1.05	43%
6	App Semester (Summer vs. Fall)	4.29	1.58	43%
7	Ethnicity	4.14	1.25	43%
8	Last School Attended	3.43	1.18	29%
9	Gender	2.86	0.99	29%
10	Zip Code	4	1.2	14%
11	City	3.71	1.28	14%
12	Distance from Pullman Campus (in miles)	3.71	1.28	14%
13	County	3.71	1.39	0%
14	Applied Day of the Week	2	0.93	0%

Narrative: The range of 24 factors accurately predicted the enrollment outcome of an admitted applicant up to 80.6% of the time. The 10 factors in table 1 were statistically significant predictors of applicant enrollment outcome while the 14 factors in table 2 had little to no predictive value.

**p < .05, **p < .01, ***p < .001*
P Value = Confidence Interval

APPENDIX C

WORKSHOP AGENDA

Workshop Agenda

- **Introduction – Approx. 10 Minutes**
    - Confirm all participants have completed the consent form; remind participants that participation is voluntary and they can opt out/leave at any time
    - Welcome everyone
    - Discuss the purpose of the research and participation in the workshop
    - Confirm everyone agrees to the workshop being recorded then being recording

- **Overview of Predictive Model & Pre-Workshop Survey – Approx. 20 Minutes**
    - Overview of the results of the predictive model
    - Overview of the results of the pre-workshop survey

- **Discussion of Predictive Model – Approx. 30 Minutes**
    - Group discussion – Reflect on the results of the model
    - Group discussion – Reflect on the results of the survey

- **Discussion of Model Application – Approx 30 Minutes**
    - Group Discussion – Can the model influence strategy & decision making? How?
    - Group Discussion – How can the model be improved?
    - Group Discussion – Can/should the model be applied to our work? How?

- **Conclusion**
    - Thank participants for participating
    - Outline next steps (post-workshop survey)
    - End recording

APPENDIX D

MODEL FOUR – VARIABLE RESULTS

Variable	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Washington Resident	0.790	0.086	84.776	1	0.000	2.203	1.862	2.607
Admitted Month								
Reference Month: July			58.241	11	0.000			
January	-1.692	0.341	24.606	1	0.000	0.184	0.094	0.359
February	-1.509	0.295	26.207	1	0.000	0.221	0.124	0.394
March	-1.337	0.254	27.636	1	0.000	0.263	0.160	0.432
April	-0.679	0.219	9.597	1	0.002	0.507	0.330	0.779
May	-0.675	0.193	12.288	1	0.000	0.509	0.349	0.743
June	-0.137	0.167	0.676	1	0.411	0.872	0.628	1.209
August	-0.011	0.149	0.005	1	0.943	0.989	0.738	1.326
September	-2.592	0.581	19.909	1	0.000	0.075	0.024	0.234
October	-2.934	0.543	29.142	1	0.000	0.053	0.018	0.154
November	-2.745	0.501	30.071	1	0.000	0.064	0.024	0.171
December	-2.176	0.430	25.559	1	0.000	0.114	0.049	0.264
Degree Name								
Reference Major: Psychology			68.569	23	0.000			
Biology	-1.060	0.209	25.705	1	0.000	0.346	0.230	0.522
Hospitality Business Management	-1.047	0.311	11.313	1	0.001	0.351	0.191	0.646
Exploring	-0.649	0.227	8.157	1	0.004	0.522	0.335	0.816
Senior Living Management	-0.623	1.063	0.344	1	0.558	0.536	0.067	4.307
Special Ed Teaching Endorsement	-0.594	1.129	0.277	1	0.599	0.552	0.060	5.046
Non-Degree	-0.493	0.925	0.284	1	0.594	0.611	0.100	3.746
Environmental & Ecosystem Sci	-0.386	0.265	2.122	1	0.145	0.680	0.405	1.143
Management	-0.314	0.141	4.955	1	0.026	0.731	0.554	0.963
Economics	-0.269	0.234	1.316	1	0.251	0.764	0.483	1.210

Marketing	-0.208	0.196	1.127	1	0.288	0.812	0.553	1.193
Data Analytics	-0.106	0.213	0.250	1	0.617	0.899	0.592	1.365
Criminal Justice	-0.047	0.183	0.066	1	0.797	0.954	0.667	1.365
Humanities	0.018	0.240	0.006	1	0.940	1.018	0.636	1.629
Social Sciences	0.033	0.202	0.026	1	0.871	1.033	0.696	1.535
Accounting	0.035	0.150	0.055	1	0.815	1.036	0.772	1.389
Anthropology	0.046	0.312	0.022	1	0.883	1.047	0.569	1.928
English	0.060	0.223	0.073	1	0.787	1.062	0.686	1.645
Sociology	0.072	0.233	0.095	1	0.758	1.075	0.680	1.698
History	0.090	0.223	0.161	1	0.688	1.094	0.706	1.694
Human Development	0.253	0.218	1.348	1	0.246	1.287	0.840	1.972
Non-Cert Post Bacc	0.318	0.292	1.191	1	0.275	1.375	0.776	2.435
Political Science	0.318	0.267	1.425	1	0.233	1.375	0.815	2.318
Integrated Strategic Communications	0.333	0.214	2.428	1	0.119	1.396	0.918	2.123
Admitted Day of the Week								
Reference Day: Monday			35.148	6	0.000			
Tuesday	0.257	0.120	4.571	1	0.033	1.293	1.022	1.638
Wednesday	-0.045	0.119	0.143	1	0.705	0.956	0.757	1.207
Thursday	-0.006	0.121	0.003	1	0.958	0.994	0.783	1.260
Friday	0.041	0.122	0.111	1	0.740	1.042	0.819	1.324
Saturday	1.055	0.247	18.292	1	0.000	2.872	1.771	4.659
Sunday	0.770	0.258	8.913	1	0.003	2.160	1.303	3.580
Degree Type								
Reference Variable: First Time Undergraduate			25.894	2	0.000			
Non-Degree Seeking	-1.650	0.969	2.901	1	0.089	0.192	0.029	1.282
Post-Baccalaureate	-1.001	0.207	23.327	1	0.000	0.368	0.245	0.552

First-generation Flag (1/0)	0.195	0.077	6.459	1	0.011	1.215	1.046	1.412
Total Transfer Credit	0.004	0.001	12.966	1	0.000	1.004	1.002	1.006
Admission Turnaround Time (days)[2]	-0.004	0.002	5.991	1	0.014	0.996	0.993	0.999
Application Month								
Reference Month: July			31.409	11	0.001			
January	0.662	0.316	4.385	1	0.036	1.938	1.043	3.600
February	0.568	0.289	3.866	1	0.049	1.764	1.002	3.106
March	0.280	0.252	1.236	1	0.266	1.323	0.807	2.169
April	0.121	0.227	0.282	1	0.595	1.128	0.723	1.760
May	0.128	0.198	0.420	1	0.517	1.137	0.772	1.674
June	0.090	0.168	0.290	1	0.590	1.095	0.788	1.521
August	1.544	0.340	20.663	1	0.000	4.683	2.407	9.114
September	1.409	0.525	7.200	1	0.007	4.092	1.462	11.455
October	1.489	0.474	9.865	1	0.002	4.431	1.750	11.219
November	0.749	0.433	2.999	1	0.083	2.115	0.906	4.938
December	1.049	0.383	7.511	1	0.006	2.853	1.348	6.040
Student Type								
Reference Type: Transfer			16.902	3	0.001			
First Year Student (Freshman)	-0.326	0.135	5.863	1	0.015	0.722	0.555	0.940
Former Student Returning	0.275	0.108	6.471	1	0.011	1.317	1.065	1.627
Non-Degree Seeking	0.151	0.423	0.128	1	0.721	1.163	0.508	2.665
Constant	0.144	0.206	0.490	1	0.484	1.155		

[2] Odds ratio of -0.004 represents the relative change in enrollment outcome for a single day change in the length of admissions turnaround time (time from apply to admit decision). In more practical and operational terms, the odds of enrolling decrease by 4% for every additional 10 days applicants wait for an admissions decision. (Variable Mean = 34.8 days)

APPENDIX E

QUALITATIVE CODES, CATEGORIES, AND THEMES

Themes	**Codes**	**Examples**
1. Desire to Understand	1.1 Reservation	“To a certain extent but statistically the data can be misleading as well.” (P6-Workshop)
		“There are non-predictive factors that are still important that aren’t statistically important so I’m not sure you cannot completely rely on the predictive model.” (P2-Workshop)
		“We need to look into the data further and understand what is influencing each of those factors.” (P4-Workshop)
		“I’m really wondering about admitted Day of the week. Now I wonder if you could share some insight into that, because it just doesn’t make sense to me.” (P3-Workshop)
		“I’m weary of the admitted day of the week variable because I don’t understand it”. (P2-Workshop)
		“When you say number of days between apps submitted and first day of classes is not predictive. But then you say that application month is predictive. It seems like a conflict I don’t understand.”(P2-Workshop)
	1.2 Seeking Clarity	“I got a little bit wrapped around the wheels, I think, because some of these things overlap and correlate strongly, but some said they were predictive, and some said they were not but they seem very similar.” (P2-Workshop)
		“Do you agree that we could have interpreted it, or had a reason for saying it was correlated that differed from one person to the next?” (P3-Workshop)
		“I had a question about one of the nonsignificant factors when you identified the application semester, summer versus Fall, which I thought was interesting. I would have expected fall versus spring. Could you talk a little bit about what that means? Maybe I’m not interpreting it correctly.” (P1-Workshop)
		“I just have a question. On the admitted day of the week variable, how was that date decided? Was it the day the staff entered the admissions decision or the day we send them the admissions letter? Or the day they accepted their admissions? Like, how did you decide what day to use?” (P4-Workshop)
		“Does enrolled mean on the first day of class, or that they enrolled in classes on census

	day?” (P2-Workshop) “Is there a reason why you went with ethnicity, but then excluded race?” (P5-Workshop) “Were international students included?” (P7-Workshop)
1.3 Wanting to Learn More	“It provides some good information, I think. I would like to know if there was any difference in yield from fall to spring?” (P1-Workshop) “Ethnicity jumped out to me. I didn’t think that would be non predictive when I was looking at the list.” (P4-Workshop) “As I was looking at this, I found myself wanting to know more about each category, like the States. Like, what is the yield for New York vs. Connecticut. Or the majors, like, which major is better, and which is worse?” (P6-Workshop) “Which month predicted the most, the highest, ability to yield?” (P2-Workshop) “I feel like I really want to dig into this data to see what is actually happening and what is influencing each variable.” (P5-Workshop)
1.4 Connecting to Local-Level Knowledge or Experience	“I would assume that, like, former students returning applicants are probably way more likely to confirm than other people, and that probably would have an influence over the data.” (P7-Workshop) “But the fact that non degree was in there kind of surprised me, because non degree has to influence all of these areas significantly”(P2-Workshop) “There is a very rare exception that we would actually admit a student on a Saturday or Sunday. So, I want to learn more why that variable was significant.” (P3-Workshop) “Well, to your point, a lot of stuff in the system has a delay. Like, say, on a Friday. But the process doesn’t run until after midnight, and so perhaps that date wouldn’t be recorded until Saturday or Sunday.” (P5-Workshop) “I wonder if age will change in the future since we are getting a larger portion of traditional-aged students applying.” (P6-Workshop)
1.5	“Yeah, thank you. That makes sense. And I think it could have accounted, perhaps, for the way we answered questions, too, how we interpreted the definition of a variable.”

	Acceptance / Understanding	(P3-Workshop) “That makes way more sense. Thank you for clarifying.” (P4-Workshop) “It makes sense now why some of these variables could be combined or eliminated.” (P2-Workshop)
2. Desire for Application	2.1 Optimism	"Just to add, I feel like this information you have really helps us ask the right questions in each category where, like it, kind of directs the conversation where we need to spend more time exploring." (P3-Workshop) "There is so much data here, but it gives us a starting place to ask deeper questions, like, why is this significant but this isn't? But this has been definitely helpful." (P6-Workshop) "This is so exciting, I just want to learn more." (P7-Workshop) "You just have to be strategic with what resources we have and how much allocation of time we have, and this could help us focus our efforts." (P1-Workshop)
	2.2 Recommendation / Suggestion	"Going forward as we're looking at future iterations, maybe we can narrow down variables that might overlap to more broad categories." (P2-Workshop) "So, I've never built a predictive model, but I read about them and look them up and stuff like that. 10 seems like a lot of predictors. Can we build a less complex model with fewer factors that is just as predictive?" (P3-Workshop)
	2.3 Opportunity to Improve Upon Model	"I think future models need to include spring in our predictive model because our campus has a large increase in spring students compared to other campuses." (P1-Workshop) "There are a lot of models out there that predict persistence. But I don't know if I've read anyone that starts at the admissions level, so that would be a very, very interesting way to expand the model." (P5-Workshop) "I get asked about cost of attendance all the time, so it would be, like, cool to include financial factors in a future model." (P7-Workshop) "Can we use this to focus on certain groups? For example, can we just explore freshmen students and build our recruitment efforts for a single population instead of

	doing it for everyone?" (P6-Workshop) "I think the transfer credit variable could be expanded from total credits to the percent of credits that were transferable, or even the number of credits that actually applied towards degree completion. I think there could be something interesting there." (P2-Workshop)
2.4 Application to Strategic Planning	"Another thought I had, and this I guess this kind of goes with like our strategic planning goals, when I saw the fact that last school attended was not on the list, I was shocked. But then I also thought like, this is definitely an opportunity for us to learn more about that category for planning purposes." (P4-Workshop) "It just made me think of, like, the goals and the conversations we've had with the strategic planning group and how we can use this to shape those conversations."(P2-Workshop) "Well, looking at what you said with the application month being one of the predictors, maybe we don't target our applicants until we get closer to summer and fall, maybe changing our application strategies." (P5-Workshop) "We need to design our systems to be able to be responsive to data. Like legacy data needs to be collected on all students." (P1-Workshop) "We can use this data to identify gaps and challenges that need addressed, like outreach to underserved audiences." (P3-Workshop) "We need to be strategic in how we use our limited resources, and this could help us focus our efforts."(P1-Workshop) "It can also drive strategy towards populations and/or degrees we would like to increase yield for and see if the efforts worked." (P6-Workshop)
2.5 Application to Decision Making	"I kept thinking, what does the number of days it takes for an applicant to be admitted mean for our communication plan? My mind started going down a rabbit hole like, do we need to redo the timing of all of our communication to be informed by this data?" (P3-Workshop) "Since applicants yield higher closer to the deadline, maybe we need to encourage early applicants to start in a sooner semester instead of waiting." (P2-Workshop)

"We should consider changing our approach to align efforts closer to when applicants yield." (P6-Workshop)

"We need to look into each one of these variables and ask, why is that the case? So, we can understand it more and use it to make decisions." (P4-Workshop)

"I want to use the data so we don't exclude groups of people or underserved audiences just because they don't historically yield." (P1-Workshop)

"We could use this info to look into each category further and consider improving processes. It can also help staff decide the best way to spend their time." (P2-Workshop)

"We are running in a limited resources situation, and we have to be strategic in how we use the resources either to change the influence of the factors we have identified or to focus the resources we have on the area where students will see the greatest success." (P7-Workshop)

APPENDIX F

IRB EXEMPTIONS

EXEMPTION GRANTED

Ruth Wylie
Division of Educational Leadership and Innovation - Tempe
480/727-5175
Ruth.Wylie@asu.edu

Dear Ruth Wylie:

On 1/23/2023 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Workshop On a Predictive Model of Online Student Enrollment Decisions
Investigator:	Ruth Wylie
IRB ID:	STUDY00017226
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Consent Letter _ IRB _ Singer.pdf, Category: Consent Form; • Global Campus - Permission to Complete STUDY00017226.pdf, Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • IRB Protocol_Singer_EDM2.docx, Category: IRB Protocol; • Supporting Documents - PostWorkshop Survey_Dissertation_Singer.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Supporting Documents - PreWorkshop Survey_Dissertation_Singer.pdf, Category: Measures (Survey questions/Interview questions /interview guides/focus group questions); • Supporting Documents - Workshop Agenda_Singer.pdf, Category: Other;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (2)(ii) Tests, surveys, interviews, or observation (low risk) on 1/20/2023.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required. Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Cody Singer
Ruth Wylie
Cody Singer
Lydia Ross

EXEMPTION GRANTED

Ruth Wylie
Division of Educational Leadership and Innovation - Tempe
480/727-5175
Ruth.Wylie@asu.edu

Dear Ruth Wylie:

On 12/22/2022 the ASU IRB reviewed the following protocol:

Type of Review:	Initial Study
Title:	Educational Data Mining: A Predictive Model of Student Enrollment Decisions of Online Learners
Investigator:	Ruth Wylie
IRB ID:	STUDY00017132
Funding:	None
Grant Title:	None
Grant ID:	None
Documents Reviewed:	• Consent Letter - WSU Global Campus, Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • Consent letter from WSU, Category: Off-site authorizations (school permission, other IRB approvals, Tribal permission etc); • IRB Protocol_Singer_EDM1.docx, Category: IRB Protocol;

The IRB determined that the protocol is considered exempt pursuant to Federal Regulations 45CFR46 (1) Educational settings, (4) Secondary research on data or specimens (no consent required) on 12/22/2022.

In conducting this protocol you are required to follow the requirements listed in the INVESTIGATOR MANUAL (HRP-103).

If any changes are made to the study, the IRB must be notified at research.integrity@asu.edu to determine if additional reviews/approvals are required.

Changes may include but not limited to revisions to data collection, survey and/or interview questions, and vulnerable populations, etc.

Sincerely,

IRB Administrator

cc: Cody Singer
Cody Singer
Ruth Wylie
Lydia Ross