Predicting At-Risk Students from Disparate Sources of Institutional Data

by

Ajay S. Rayasam

Bachelor of Science in Business Administration, Babson College (2012)

Submitted to the Integrated Design and Management Program
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Author .....
Integrated Design and Management Program
May 21, 2020

Certified by .....
Matthew S. Kressy
Executive Director, Integrated Design and Management Program
Thesis Supervisor

Accepted by .....
Steven D. Eppinger
Faculty Co-Director, Integrated Design and Management Program

Predicting At-Risk Students from Disparate Sources of Institutional Data

by

Ajay S. Rayasam

Submitted to the Integrated Design and Management Program
on May 21, 2020, in partial fulfillment of the
requirements for the degree of
MASTER OF SCIENCE IN ENGINEERING AND MANAGEMENT

Abstract

In the past few years, the Mental Health Crisis in Higher Education has captivated the nation. This may be due in part to high profile cases, shifts in cultural attitudes, or increased demand for treatment. Regardless of the cause, student mental health has now become an epidemic. At MIT, there are over 4,000 consultations, 200 well-being checks and 50-70 psychiatric hospitalizations annually. In order to combat this challenge, most institutions invest in services such as mental health counseling or emergency response teams. However, these services are primarily used for students who self-report symptoms or for extreme cases. Unfortunately, of the nearly 3 million college dropouts per year, more than 40% did not report their mental illness.

While the institutions have promoted mental health awareness, many students, who suffer from mental illness, remain undiscovered. As a result, this thesis proposes an novel approach — using artificial intelligence to identify those hidden students. By leveraging non-invasive data found within the institution, machine learning can predict at-risk students before any symptoms occur. By doing so, the institutions could prevent dropouts, leaves of absences and deaths due to mental illness.

Thesis Supervisor: Matthew S. Kressy
Title: Executive Director, Integrated Design and Management Program

Acknowledgments

This work represents the culmination of a three-year personal and professional journey. And, I would be remiss if I said that I had done this alone.

First, I'd like to thank Matt Kressy. Three years ago, you took a chance on me and, by doing so, you fulfilled my childhood dream. Words cannnot describe how grateful I am. Thank you for everything.

To all those from IDM, thank you for being a part of my life. To Andy MacInnis, thank you for believing in me before I could believe in myself. To Tony Hu, thank you for always being a positive sounding board. To Melissa Parrillo, thank you for taking a chance on me. To Lesley Britton, you convinced me to apply and it changed my life. To Grace Agosto, thank you for your calm presence, infinite patience and for always fitting me into Matt's schedule. To Shandel Williams and Laura Barajas, thank you for always taking the time to hang out with me.

To MIT and the administration, thank you for providing me with this opportunity. To David Randall, I'm humbled by your generosity. To Jason Lloyd, your optimism is what kept me going through all those difficult moments. To Brian Canavan, without you, I *literally* would not have graduated on time.

To my family, thank you for providing me with unconditional love and for always supporting my dreams. To Sara Gong, thank you for helping me become the best version of myself. Without you and your support I would not have completed this thesis. If we can survive the coronavirus, we can survive anything.

To all my friends, you all are extraordinary human beings. I've been so blessed and fortunate to have met so many people who've had a profound impact in my life. Without each and everyone of you, I would not be the person I am today.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

My name is Ajay. I'm a graduate student at MIT, I have two fiercely loving parents, and a brother and sister-in-law who recently opened a restaurant. I've led business development at a startup, worked as product manager at IBM Design, and secured not only one but two internships in venture capital. By all accounts, I've lived a good life and have a bright future.

Despite all of this, I've spent most of life struggling with my self-esteem, self-worth, and self-confidence. That's because I've suffered from anxiety, mild depression and attention deficit disorder. Every day, I smile, laugh and talk to as many people as I can. If you've seen me around campus, chances are high that we've had a random conversation and, somehow, we have a few mutual friends. For most people I interact with, they would never guess. For my inner circle, they've seen the mood swings, tantrums, and moments of reprieve. In my opinion, mental illness is the true invisible enemy of our society.

I've spent most of my life unaware and in denial, but I'm sharing my story now because I'm no longer afraid. More importantly, I believe that it could make a difference in someone's life. As a kid, I struggled to read. At twenty one, I was legally blind. At MIT, I felt like an imposter. Each fact or opinion reinforced a negative feedback loop. I responded by prioritizing my career at all costs. And, I

missed out on a lot of things because of it. I never went on spring break, completed a bar crawl, or even allowed myself to get too close to others.

This vicious cycle would've never ended if I didn't attend MIT. Through casual conversations during my first semester, I became aware that I had reached my breaking point — I couldn't read a book, attend class or even think about a career anymore. I needed help and my friends inspired me to seek treatment. Truthfully, therapy changed my life. It's like hiring a personal trainer, but for your mind. By nature, we're irrational and emotional beings. I no longer denied that part of me. Instead, I embraced it and took a medical leave of absence.

Prior to grad school, I underwent an experimental surgery to implant plastic inserts into my cornea. Although it enabled me to see again, I still suffered from the underlying trauma that losing your vision can cause. The gap year allowed my mind to heal. Through many therapy sessions, I also discovered that I had suffered from learning disabilities. Essentially, for my entire life, I was working twice as hard as those without a disability and it produced an enormous amount of anxiety. Instead of relaxing or spending time with friends, I would stress out over a deadline. Then, I would become depressed over the endless 'fomo'. If I didn't take the initiative to make a change, I wouldn't be here today. Now, I no longer worry about the what-ifs, comparing myself with others, or making everything perfect. I'm free to focus on what I choose. Although I'm uncertain of my path forward, I'm confident that I'll be able to figure it out.

The reason I chose this topic is because I want other students to feel the same way. I'm privileged to have the courage, support and resources to make this happen. However, most aren't. So, to the student who is suffering from anxiety, depression, imposter syndrome, learning disabilities or an eye condition, I want you to know that you're not alone. I wrote this thesis for you and my goal is to find a way to help.

Chapter 2

Background

2.1 What is Mental Health?

Mental health is an all encompassing word that may elicit a range of emotions. It's important to not only understand the definition, but also to grasp the subtle nuances of the topic. In the past century, our attitudes and approach towards mental health have changed dramatically. This shift has had a profound impact on college and university campuses across the nation.

By 2030, the World Health Organization (WHO) has forecasted that the largest health risk on earth will be depression. Despite our unprecedented abundance, depression, anxiety disorders, addiction and suicide are skyrocketing within our society. All of these conditions severely affect our mental health. According to the Center for Disease Control and Prevention (CDC), mental health is the level of a psychological well-being or an absence of mental illness. It is a state in which someone is functioning at a satisfactory level of emotional and behavioral adjustment [7].

Like our physical health, we can make a determination of whether our mental health is in a poor, average or good state. When we are in good mental health, we

can enjoy life, creating a balance of life and work activities to achieve psychological resilience, and vice versa. In other words, when an individual is in good mental health, he or she is more likely to increase activities that have a positive impact on overall health. If the individual is not in good mental health, he or she is more likely to participate in consequential activities like smoking. Studies have found that depression has been linked to a 67% increase in a person's risk of dying from heart disease [4].

For this work, it's important to understand the difference between mental health and mental illness. Mental health refers to the state of mental well-being whereas mental illness refers to an illness or disease that affects the way people think, feel, behave or interact with others. A person can experience poor mental health, but not be diagnosed with a mental illness. Likewise, a person diagnosed with a mental illness can experience periods of positive mental well-being. As forecasted by the WHO, mental illness is among the most common health conditions in the United States. More than 50% of US citizens will be diagnosed with a mental illness or disorder at some point in their lifetime [7]. While there isn't a single cause for mental illness, a number of factors, such as genetics or adverse life experiences, may contribute.

The most common mental illnesses are anxiety disorders and depression, which constitute over 25% of the United States population [3]. Anxiety disorders are characterized by feelings of worry or fear that are strong enough to interfere with one's daily activities. Types include panic attacks, obsessive-compulsive disorder, or post-traumatic stress disorder. Depression is a mood disorder in which there is significant loss of interest in activities causing significant impairment in daily life. I mention these particular mental illnesses because they are the most common among college students in the nation. Unfortunately, most mental illnesses have their peak onset during young adulthood. By the age of 25, 75%, of those who will be diagnosed, have had their first onset. Statistically, over 11.9% of college students suffer from anxiety disorders and 9% suffer from depression [39]. Together, over 20% of all mental illnesses can be traced to anxiety and depression.

2.2 The 'Mental Health Crisis in Higher Education'

The 'College Mental Health Crisis' refers to the rising number of college students who experience mental health issues and the overall demand for mental health services on college campuses. However, through my research, I believe that the crisis is much more nuanced. While demand is a critical component, access to care and a lack of trust are culprits as well. This confluence of factors have directly led to the mental health crisis in higher education.

2.2.1 Demand for Mental Health Services

The key driver for this epidemic has been the demand for mental health services. Between 2009 and 2015, the number of students visiting counseling centers have increased by about 30% on average, while enrollment grew by less than 6% [29]. Data from the Healthy Minds Study, an annual mental health survey that tracked responses from over 150,000 students, further supports these trends. From 2007 to 2019, the rates of students who sought therapy have increased from 13.3% to 29.9%. The rate of students who used any psychotropic medication increased from 11.8% to 23.9%. The proportion of students with a diagnosed mental illness increased from 21.9% to 35.5% and the most common mental illnesses are substance use, anxiety, and depression [2].

There are many potential causes for this demand. College is a stressful time for students as they juggle classes, work, independent living, and career aspirations. It's often the first time they no longer have their parents to keep them on track. On a macroeconomic level, there's been an unprecedented rise in student debt compounded by multiple recessions. There's also the new hyper-connected world of social media, which provides an endless array of choices that may overwhelm and confuse students. While these factors have potential influence, the more likely root cause is that the onset of over 75% of mental illnesses occurs before the age of 25.

2.2.2 Lack of Access to Care

This generation of students has been known to destigmatize mental illness, making it more likely that students will seek help. As a result, access to care is harder than ever. Unfortunately, university and college counseling centers were not designed for this volume. It is estimated that 20-35% of college students need mental health services, but the average counseling center utilization rate is 11.8%. Since counseling centers are assumed to be running at full capacity, this means that almost 8-13% of students are not being served by those centers [18]. In fact, only 46% of centers in the United States offer psychiatric services on their campus. Of those centers, 57% reported that they need more hours of psychiatric services than they currently have. For centers that have a waitlist, the average wait was 17.7 business days. However, these centers have a limited ability to increase capacity since the average budget is $967,000 [31].

To make things worse, the university centers primarily refer students to off-campus mental health professionals. This process is convoluted. First, the student must have health insurance and, in many cases, use their parent's insurance. Second, the student should provide a list of preferences, such as gender, ethnicity, schedule, type of professions or treatment style. This assumes the student may have prior knowledge, which is most likely not the case. Third, the mental health professional must accept that information and have availability to provide treatment. In a sense, finding treatment is a complex optimization algorithm with many constraints. Although there are case managers to assist, there are not enough to adequately manage the students. For example, at MIT, there is one referral specialist to serve over 10,000 students.

Adding to the distress, there is a shortage of psychologists and psychiatrists [8]. In the US, there are approximately 30,451 practicing psychiatrists, which equates to about 9 psychiatrists per 100,000 people, short of the nearly 15 per 100,000 recommended to provide good mental health care. This shortage creates immense competition for treatment. Typically, a psychiatrist, a medical doctor who specializes in

mental health and can prescribe medication, charges an average of $100-300 per hour. Many psychiatrists are sole practitioners who charge cash payments, which are considered out-of-pocket expenses. For those who can afford the payments, they may be able to find treatment. For the rest, students are competing with other populations for a limited supply of mental health professionals. Unfortunately, only 55% of all psychiatrists, in the United States, accept insurance [15].

The other 45% of psychiatrists don't accept health insurance because the pay rate, from health insurance providers, are far below their personal rates. Researchers reported that private insurance companies are paying 13% to 14% less for mental health care than Medicare does [40]. If a provider finds a reason to cancel pay for a treatment, the psychiatrist can't bill the patient separately and the visit becomes a sunk cost. Therefore, psychiatrists don't have a monetary incentive to accept insurances. Unfortunately, this isn't just a student population problem. An office visit with a therapist is 5 times as likely to be out-of-network and is more expensive than a primary care appointment [27]. Of the 45.6 million American adults living with mental illness, fewer than half received treatment and the most common reason for not seeking treatment was that patients couldn't afford it [25].

2.2.3 Lack of Healthcare Coverage

This last paragraph introduced a daunting aspect of this crisis. Fundamentally, health insurance coverage for behavioral health is inadequate because of misaligned incentives. Many insurance companies are for-profit enterprises that collect a monthly payment, called a premium, in exchange for promising to pay a specific amount of money for a specific asset loss, such as someone's life and forgone earning potential. Essentially, for-profit insurance companies want to pay as little as possible. The more mental health coverage, the lower the profit margin. To cover expenses, the insurance companies will charge higher premiums and lower the value of coverage for those patients with pre-existing conditions.

As a result, insurance companies will go to great lengths to avoid paying. For example, there is evidence of the use of "ghost" or phantom networks. To provide some context, an insurance company provides a list of mental health professionals who accept their insurance. A ghost network is a list that is intentionally filled with out-of-date and inaccurate contact information. In a recent study, researchers called 360 psychiatrists on Blue Cross Blue Shield's in-network provider lists in Houston, Chicago, and Boston. When the researchers actually reached psychiatrists' offices, many of the doctors didn't accept the insurance or weren't taking new patients. After calling every number twice, the researchers were unable to make appointments with 74% of providers on the list [34].

The existence of these ghost networks have led to several lawsuits. In 2019, a federal judge ruled that UnitedHealth was systematically and illegally gaming the system to deny mental health care to its insured customers in order to improve the company's profitability. In 2018, Aetna settled a lawsuit in Massachusetts after the attorney general found that its provider directories are inaccurate and deceptive [42].

However, there is some room for optimism. In 2016, the Department of Health and Human Services began shifting to value-based care. Value-based care is a form of reimbursement that ties payments for care delivery to the quality of care provided. In the past, reimbursement occurred retrospectively for services delivered based on bill charges or annual fee schedules. The payments reward healthcare providers for both efficiency and effectiveness. Some performance goals include patient outcomes, reducing hospital re-admissions, using certified health IT, and improving preventative care.

As of 2020, 59% of healthcare payments are tied to value-based care [35]. This payment model incentivizes hospitals to reduce their burden by rewarding proactive measures. With more progress, the model could be modified to include preventative behavioral health measures. Although other recent developments, such as increased tele-therapy coverage due to the Covid-19 pandemic, have also provided more hope,

the long-term outlook for reimbursing mental healthcare is uncertain.

2.2.4 Lack of Trust

While the demand, the limited supply of professionals and healthcare itself has contributed to the crisis, the heart of the issue is the lack of trust between the students and the administrations. At times, this relationship has been contentious at best. The main point of contention is the use of the involuntary withdrawal policy. The policy states that the institution has the right to involuntarily withdraw a student in particular situations. The student must pose a credible risk of substantial harm to a member or members of the university community (which could include self-harm) or substantially impede the educational, residential, or other activities of the university community [23].

In some cases, the wording of this policy can circumvent the Americans with Disabilities Act and Section 504 of the Rehabilitation Act, which requires schools to provide children with mental and physical disabilities the same access to educational programs and services that other children have [30]. In other words, students are afraid to report their mental illness because the institution could expel them.

Although there is no proof that institutions have done this in the past, a landmark case will hopefully spark change within the industry. *Mental Health and Wellness Coalition v. Stanford* was a lawsuit that alleged that Stanford University repeatedly violated state and federal anti-discrimination laws in its response to students with mental health disabilities, including those who have been hospitalized for suicide attempts. In 2019, a historic settlement was reached, although not an admission of liability, in which Stanford agreed to revise its involuntary leave of absence policy. Under the revised policy, Stanford will give students the option of taking a voluntary leave and "will not discriminate against students with mental health disabilities" who choose to do so.

It remains to be seen what impact this case will have as more than 50% of post-secondary institutions still have an involuntary withdrawal policy [20]. It's also unclear whether other schools have updated their policies in response to the case. What is certain is that the mental health crisis in education is a complex issue with many potential causes. Some, like healthcare, remain out of the control of the students and administrations. For those items within control, there needs to be improvement and continuity. Otherwise, the mental health crisis in higher education may never be resolved.

2.3 A Matter of Life and Death

While it's easy to get mired in the complexities of this topic, it's more important to remind ourselves why we are so concerned with mental health and mental illness. It's not just about helping students reach their full potential, but it's also about preventing the worst scenario. Suicide is the 10th-leading cause of death in the United States, but is second-most among college-aged students. The rate of suicidal attempts have increased by 2.6%, peaking at 10.6% of the student population this past year [10].

The attention has increased exponentially due to high profile cases. In 1998, Philip Gale committed suicide by jumping from Building 54 at MIT. Prior to his action, he attempted to seek medication from a psychiatrist and friends, citing boredom and depression [33]. In 2015, Luke Tang committed suicide in his Harvard dormitory despite signing a contract promising to follow a treatment plan [36]. In 2019, 9 students committed suicide in a single semester at USC [9]. Sadly, these deaths are not limited to the students themselves. Last year, Gregory Eeels, a highly regarded director of psychological services at UPenn, committed suicide. He had complained about the demands and stress of his new job, and how it kept him from his wife and three children.

Although it's uncomfortable to discuss these tragedies, I mention them because

they all have one thing in common — their deaths were preventable. A majority of Americans agree that people who are suicidal can be treated and go on to live successful lives (91%) and that suicide can often be prevented (87%). By speaking openly, we can all do our part to help those in need.

2.4 Mental Health and Student Retention

Suicides tend to grab our attention, but the vast majority of students remain imprisoned by their mental illness. Though we can measure the rate of suicide, it's more difficult to estimate the impact on those who are still alive. A reasonable method is to determine the potential loss of productivity due to mental illness and its macroeconomic consequences on our society.

When considering the potential loss of productivity for students, prototypical measurements like grade point average (gpa) or post-graduate employment status come to mind. In particular, high school gpa and first semester gpa are significant predictors of student retention, which is the percent of students who return to the same institution [43]. In fact, 45% of college dropouts that occur in the first two years of college can be attributed to what students learn about their academic performance [44]. Poor academic performance is also a leading indicator of graduation rates.

From an institution's perspective, poor student retention leads to lower graduation rates, which produce fewer tuition dollars and inferior national rankings. As a result, institutions place an emphasis freshmen experience because most post-secondary dropouts occur during the first year [47]. From a student's perspective, there may not be a correlation between gpa, employment or lifetime earnings. However, there is a significant lifetime earnings gap between high school graduates and college graduates. Therefore, we can assume that a loss in productivity can lead to a lower gpa, which can influence a college student to dropout. By dropping out, the student may decrease his or her lifetime earning potential.

As alluded to, the viability of higher education institutions is predicated upon student retention. Collectively, these institutions have invested in solutions to address known causes such as financial burden, family issues, physical ailments, or poor academic preparedness. However, the freshman persistence rate, the percent of students who return to college for their second year, has barely improved. Of the 3.5 million students who enrolled in college for the first time in fall 2017, 74 or 2.6 million students persisted as of fall 2018, representing only a 2% increase from nearly a decade ago [11]. This data indicates that higher education leaders, scholars and administrators are overlooking the impact of mental illness on student retention.

In the past few years, there has been more evidence that suggests a correlation between mental illness, academic performance and student retention. The effect of mental illness upon gpa is substantial and immediate. Those who are formally diagnosed with depression were associated with a -0.49 point, or half a letter grade, decrease in gpa [26]. Specifically, students with mild to severe depression had an average -0.2 change in gpa the semester of onset, and if there is co-occurring anxiety, that figure becomes -0.4 [46]. This reduction in academic performance leads to lower student retention.

According to Healthy Minds Study, students with mental health problems were twice as likely to leave an institution. The College Life Study also found that mental health problems predicted a gap of enrollment of one or more semesters [12]. It discovered that students who were diagnosed with depression were more likely to drop out. And, a longitudinal study determined that students, who had a low gpa and mental health problems, were more likely to drop out than students with just a low gpa. Since many institutions use gpa to identify students who may drop out, this study also confirms that including mental health data would identify 19% more at risk students [22]. In sum, this research supports a potential link between mental health, academic performance, persistence, and retention.

2.4.1 Economic Impact on Students

The impact of the mental health crisis is not just limited to statistics, correlations or deaths. There are far reaching consequences that have a macroeconomic impact on the future. Based on data from 2010, the global direct and indirect economic costs of mental disorders were estimated to be $2.5 trillion [22]. The indirect costs, which include income loss due to disability and estimated productivity losses, were estimated to be $1.7 trillion are much higher than the direct costs.

According to the Department of Education, the 6-year graduation rate for full-time undergraduate students who began seeking a bachelor's degree at 4-year degree-granting institutions was 62% [1]. Since there are about 14 million college students in the United States, 5.32 million are expected to drop out. Of those dropouts, 3.5 million students, or approximately 64%, left college due to mental illness [24]. In addition, nearly 50 % of those students did not access mental health services [24]. Therefore, we can presume that 1.1 million dropouts, or approximately 21%, did not seek help on-campus prior to dropping out. Unfortunately, education is a key determinant of future employment and income prospects of young people. Adults with bachelor's degrees can earn approximately $900,000 more in median lifetime earnings than high school graduates [17]. **This translates into approximately $3 trillion in lost lifetime earnings for those who drop out of college due to mental illness.**

2.4.2 Economic Impact on Institutions

When reviewing the economic impact of mental illness for college students, it's important to also consider the effect on institutions. In 2016–17, total revenues at degree-granting post-secondary institutions in the United States were $649 billion. Total revenues were $391 billion at public institutions, $243 billion at private non-profit institutions, and $16 billion at private for-profit institutions [1].

For a private nonprofit institution like MIT, 30% of revenues come from tuition, 20% from investments (endowment) and 11% from government and other grants. Per full-time-equivalent (FTE) student, revenues from tuition and fees were 25% higher in 2016–17 than in 2010–11 at public institutions ($7,700 vs. $6,100) and 7% higher at private nonprofit institutions ($21,900 vs. $20,500). At private for-profit institutions, revenues from tuition and fees were 4% lower in 2016–17 than in 2010–11 ($16,500 vs. $17,100)[1]. The money from various revenues are used to fund the institutions' operations. In 2016–17, instruction expenses per FTE student was the largest expense category at public institutions ($10,800) and private nonprofit institutions ($18,400).

This means that when a large number of students drop out due to mental illness, the institutions have less money to pay for research, professors, and student support systems. If we assume that the proportion of dropout rates are normally distributed across all types of institutions, **then the total tuition revenues lost from 3.5 million students, who dropout due to mental health, is $35 billion per year.**

2.5 Current Treatment Model

2.5.1 What is Reactive Care?

Reactive care is essentially a system that waits for a problem to arise before jumping into action. By design, most healthcare organizations in the United States operate this way and higher education is no exception. While university and college counseling centers receive a lot of intention and blame, there is also a complex web of student support services that should share the burden of responsibility.

The university and college counseling centers are the primary centers for treating mental illness on-campus. The centers are organized into 3 primary models of delivering services — Absorption (37.3% of centers), Standard Treatment A (24.3% of centers), and Standard Treatment B (38.4% of centers) [31]. The Absorption model

is where the clinical staff are expected to evaluate and then assume primary clinical responsibility for a specific number of new patients each week, regardless of how many patients they were currently responsible for. Standard Treatment A is a model in which clinical staff are expected to evaluate a specific number of new patients per week, but were not expected to assume primary clinical responsibility. Last, Standard Treatment B is where clinical staff are expected to accept a new client for an initial assessment and subsequently assume primary clinical responsibility for a client only if there is an available space on their schedule. It's unclear which model is the most effective, but most schools have focused on maximizing the slots available for unique patients, so that wait times are reduced. The centers don't actively seek new patients (students), but rather treat them on a first come, first served basis. The student must actively schedule an appointment. On occasion, another student support organization will refer a student over or mention at-risk students at weekly, multidisciplinary meetings.

The centers are typically not the first point of contact for a student to seek mental health treatment. Instead, one of the many services provided by the student affairs (or student support) department is more likely to be. The student affairs department is responsible for student success at institutions of higher education to enhance student growth and development. Their primary focus is the development of the student as a "whole person". The services include all functions pertaining to academic services, admissions, financial aid, alumni outreach, campus life, counselling, health, and wellness, career services, residence programs, athletics and student conduct. Essentially, the student affairs is the hub of all things student related. As the hub, administrators within these departments often interact with students and are the first to witness signs of mental illness. In many cases, students conflate academic support services with mental health services.

2.5.2 Unintended Consequences

The rapid increase in post-secondary institutions has created an intense need to support more students. However, these services were not designed for this type of volume. With the influx of demand, the student affairs department must now become more collaborative and communicate more effectively. Otherwise, many students may fall through the cracks.

To illustrate this problem, I'm going to provide a few scenarios. In scenario 1, a student schedules an appointment with the student services team to receive an extension for a problem set. During the conversation, the students mention that they've been feeling depressed lately. The administrator has a legal obligation to report the incident. He or she may recommend mental health counseling and ask permission to make the introduction. After sending an email, the counseling center confirms the receipt, but the student never follows up. In scenario 2, another student was referred to the counseling center by the student support services. This time, the student is able to book an appointment and shows up despite waiting over a month. Since the student is a Master's candidate and the condition seems mild, the student was recommended to visit an off-campus therapist.

After several emails with the referral specialist, the student is able to find a therapist that fits a specific criteria, but another month has gone by. Unfortunately, the student fails to connect with the therapist. Instead of emailing the referral specialist and waiting another few weeks, the student drops treatment altogether. In scenario 3, a professor notices that a student has been absent for a few classes. The professor calls the director of student well-being to let them know. The director contacts the academic services department to follow up. Unfortunately, it's the end of the semester, the peak of demand has flooded the offices and the team sends an email to the graduate resident advisor to follow up, but they don't hear from the student until the next semester.

In each case, we describe a complex situation that requires tight coordination and communication. When looking at the whole picture, there are hundreds to thousands of requests in a given semester. It's easy to lose track of students due to the sheer volume of incidents. The scenarios also present another conundrum — who bears responsibility over managing the student? Is the last contact responsible? Is the student affairs department head responsible? The answer is outlined by each individual school and its internal policies. As a whole, higher education is mostly decentralized, resistant to change and risk averse. For example, each department may have a preference for a specific software or a format of communication. Some may like a cloud-based case conduct software whereas others prefer an on-premise FileMaker system. The presented scenarios also expose the underlying challenge of treating students. Students, by law, are consenting adults who have the freedom to determine whether he or she needs treatment. There are extenuating circumstances in which the student may present a danger to themselves or others, but, for the most part, the services must rely on students to proactively reach out to administrators and openly discuss their challenges.

2.5.3 Intermediate Solutions

Though higher education institutions are historically risk averse, there have been improvements in the past 10 years. First, the institutions have recognized the need and have begun to invest in mental health. From 2011 to 2018, the average budget for a university and college counseling center increased from $714,546 to $967,165 or a 35% increase [13]. The operating budget also increased from $63,017 to $92,159 or a 46% increase. There is also evidence that university and college presidents are beginning to prioritize it. 72% of the presidents indicated they had spent more money on mental health initiatives than they did three years ago [45].

The most important change has been the implementation of case managers. As the direct response to the Virginia Tech shootings in 2007, case management teams

were created to coordinate the prevention, intervention, and support efforts across campus and community systems. They are responsible for assisting at risk students and students facing crises, life traumas, and other barriers that impede success. Tasks include arranging for appropriate medical or mental health care, evaluating threats, or maintaining contact with the student. However, the responsibilities once again depend on the institution. In some schools, case managers are also responsible for identifying resources for academically struggling students. For the most part, case managers relieve the burden of other administrators within student affairs.

Unfortunately, there are still many limitations. For instance, the average number of case managers at an institution is 2, which translates into 1 case manager for every 7,115 students [20]. As a result, case managers are often forced to wear multiple hats while working with students. There are conflicts of interest and tasks that may be out of the scope of the job. In some instances, departments may skirt their responsibilities, which expands the case manager's workload. At any point, a case manager can expect to have 40 open cases and be referred to over 300 students in a given year, leading to immense stress and high turnover. Since the job requires accurate documentation and record-keeping, technology could alleviate some concerns. Unfortunately, the average operating budget is approximately $16,500, which is primarily used for marketing, professional development and supplies [20]. When you factor an average salary of $60,000, the investment from an institution averages to be around $136,500. This proves that there is still a need for more investment.

2.6 Making the Case for Proactive Care

Benjamin Franklin once said, "an ounce of prevention is worth a pound of cure." His expression meant that, when dealing with a problem, spending a small amount of time and effort upfront can save you from more trouble in the end. In the same light, institutions in higher education can take a similar approach by investing in proactive

care. By doing so, the institutions could reduce the burden on their mental health infrastructure while achieving a positive return on investment.

Proactive, or preventative, care is the idea that early intervention is more effective than reacting to an illness. To substantiate this claim, research has demonstrated that preventative measures can help alleviate mental illness. In Australia, a team proved that early intervention programs, such as increased cognitive behavioral therapy (cbt) sessions, produced positive outcomes [38]. Another study discovered that brief motivational interventions prevented alcohol misuse and that a social marketing campaign reduced some symptoms of depression or anxiety [41]. A third study suggested that technology-delivered interventions can also reduce symptoms related to depression, anxiety, and stress [19]. And, a fourth study further confirmed that identifying students with symptoms of depression and intervening early with cbt sessions were effective [16].

While the efficacy of the interventions are important, it's not enough to move the needle. Like in any other industry, higher education administrators must determine the need and calculate the return on investment, but a well defined method doesn't exist. By combining a few resources, the new method includes both short-term and long-term benefits. Short-term returns include tuition fees and cost savings. In the long-term, institutions can expect improvements in rankings, which will correlate to more applicants and higher alumni giving rates.

2.6.1 Calculating the Return on Investment

The best way to calculate the return on investment of proactive care is through an example. In a hypothetical scenario, let's assume that an institution has 10,000 students, a 20% dropout rate, and an average annual tuition of $20,000. We'll also assume that the institution will invest in preventative treatment for each dropout affected by mental illness. The estimated cost of treatment per student is $1,000 and

the estimated rate of effectiveness is 10% [21]. Of the 2000 dropouts, 64% or 1280 students suffer from mental illness. If the institution invests in treatment for each of them, they can expect to prevent 128 students from dropping out, saving $2.56 million tuition fees.

However, this calculation doesn't factor in cost savings from recruiting additional students. For decades, nonprofit colleges and universities spent around 2% of their tuition revenue on recruitment. However, with rising competition from online and for-profit schools, annual recruiting spend has skyrocketed to over $10 billion within the industry [28]. As of 2018, the average cost of recruiting a single undergraduate student is $2,357, which includes marketing expenses such as digital ads and admissions events [6]. By factoring this figure into our hypothetical scenario, the institution would save an additional $301,696 in marketing expenses.

Another calculation that could be considered in is the expected return from alumni. Specifically, about 10% students will donate an annual median gift of $1,004 [5]. Assuming a 40-year career, this will provide an additional $522,080 in funding over the lifetime of the students.

Altogether, the hypothetical institution would retain approximately $2.86 million (excluding alumni donations) while only spending $1.28 million, translating into a 123% return on investment within one year. When factoring in long-term benefits, the amount could further increase. **In sum, its cheaper for a college to retain a student than to recruit a new one.**

2.6.2 Data-as-a-Solution

While there is evidence to prove the feasibility and viability of proactive mental health care, it remains to be seen what type of intervention is most desirable. As such, there are a number of approaches to choose from, including motivational interventions, health promotion and wellness services, marketing campaigns, or brief treatments. Even so, institutions are hesitant to invest in them. This may be due in part to the lack of scalability and efficacy. For example, motivational speeches, wellness services or additional treatment require an exponential increase in staff, which may lead to more bureaucracy, liability or human error. When considering the optimal solution for mental health, the institutions should take a page out of their own playbook.

Although this research has prioritized mental health, higher education institutions have also suffered from declining graduation rates due to poor academic performance and financial stress. To combat this issue, lower ranked institutions found a cost-effective, yet scalable solution — predictive analytics. The idea is to find trends and patterns in large amounts of historical data and use those patterns to predict the future. Put another way, past behavior of former students generates predictions for current students.

For example, a student may be pursuing an academic career, but is struggling to complete a few mathematics courses. Unfortunately, thousands of students, who were accepted into PhD programs, performed well in these classes and it's a leading indicator. As a result, a predictive system would flag this student and request an academic intervention. This type of pattern analysis allows colleges to uncover which students are off course and intervene when there's still plenty of time left.

The adoption of predictive analytics was primarily the result of pressure from the Obama administration and the influx of grants from foundations. After the 2008 recession, states reduced funding for public universities and demanded higher graduation rates. Around the same time, philanthropic foundations called upon colleges to

track and measure student progress in order to keep up with other countries. Notably, the Bill and Melinda Gates Foundation provided grants to purchase data tools and software. Driving the market direction are macroeconomic factors from the Great Recession. After 2008, the American fertility rate has decreased to their lowest point in almost 40 years. As a result, the college population will decline by 15% after 2025 [14]. With a lower expected pool of applicants, schools are now prioritizing retention.

So far, an estimated 1,400 colleges and universities have implemented predictive analytics and it may be working. For the past few years, the national college graduation rates have continued rising. While there are many positives, there are an equal number of concerns relating to privacy. For example, in 2018, Georgia State began tracking how often each student connects with campus WiFi, logs into the school's computer system, visits the library and pays tuition in a timely manner [37]. Another example, the University of Arizona is tracking freshman students' ID card swipes [32]. These examples leave room for debate on the ethics of using sensitive student data to train predictive models, which may also further reinforce racial inequalities. Regardless of the concerns, institutions are paying over $300,000 per year for predictive analytics software. If they can justify this spending, they can also consider leveraging predictive analytics for student mental health.

2.6.3 Artificial Intelligence

Artificial intelligence (AI) is the technology behind predictive analytics. AI includes a broad range of use cases that include having conversations, identifying objects in photos or transcribing audio, among others. At a high level, AI is a branch of computer science that seeks to build machines that carry out tasks which, when performed by humans, require intelligence. The machines are entities that are able to receive inputs from the environment, interpret and learn from such inputs and then perform an action to achieve a particular goal or objective.

For this thesis, we'll be using a machine learning technique to prove the hypothesis. Machine learning is a technique of AI to provide insights from data. There are two methods — supervised and unsupervised. In supervised learning, a human must train the machine to identify the data. The algorithm learns from a labeled training data and can predict outcomes for unforeseen data. Unsupervised learning is a machine learning technique in which the model works on its own to discover information. It mainly deals with the unlabelled data and performs more complex processing tasks compared to supervised learning.

Collectively, machine learning algorithms are a key component of predictive analytics. You can think of predictive analytics as a blanket term that summarizes the entire process of extracting insights and building models in order meet an objective.

Chapter 3

Journey

During my gap year, I was constantly contemplating whether I should explore mental health for my thesis. Up to that point, my mental health journey had been painful and lonely. In particular, I remember my experience with MIT Mental Health. It took me 3 weeks to receive an appointment. Then, I was referred to a therapist and it wasn't a good fit. If I didn't take the year off, I probably would have stopped seeking treatment.

However, my thought process towards mental health changed when I began to think about others. Before I conducted the bulk of this research, I had to know if other MIT students felt the same way as I did. Did they feel like an imposter? Did they suffer from mental illness? Did they have a poor experience with on-campus services too?

3.1 Inspiration

One day, I finally had the courage to send a survey. Within a few hours, 65 people responded. Instead of describing their thoughts, I'd rather let the students speak for themselves:

> *"I've stopped notifying anyone because I was struggling more than ever and felt like nothing was helping. Nobody bothered to even contact me, which is terrible since they're most likely to get suicidal"- Student 1*
>
> *"People usually don't reach out to schedule their first appointment until things are already bad. But I wasn't willing to let MIT Mental Health know how bad things were because I'd heard stories about forced leaves and such" - Student 2*
>
> *"It took two more weeks to get a few recommendations, and a week more to get a response from one of them, which was far away from campus. The whole process was very inconvenient and I can easily see how people desperately in need of help can slip through the cracks" — Student 3*
>
> *"It took more than two months to get an appointment at MIT. I showed up to my appointment three days later only for them to tell me I was not even on the system. At that point, I just started crying." - Student 4*
>
> *"It was very helpful and gave me good perspective that I could take going forward. The main issue was the follow-up. At a certain point I figured the pain of trying to get something rescheduled wasn't worth it" — Student 5*

3.2 Call to Action

When I reflect on those interactions, I have trouble putting my emotions into words. It was gut wrenching to hear their stories, but it confirmed my worst fears. Students were not receiving the care that they need when they needed it. At the same time, counseling centers are understaffed and have limited resources, so I understand both sides.

Personally, those students inspired me to do something about it. They also made me feel like I wasn't alone. They gave me the courage to openly express my pain as much as my joy. With them in mind, I had to find a way to connect with the administration. So, I used the survey data to tell their story. Instead of selling the administration on a solution, my goal was to understand their needs. If I could earn their trust, I would be one step closer to helping other students.

3.3 Missing Link: Student Support Data

Over the past 6 months, I've met with all levels of the administration, from IST to the Registrar. Eventually, I earned the support of David Randall, the Senior Associate Dean of Student Support and Well-being. David is responsible for five departments within the Division of Student Life, including student support services, care & response, alcohol & other drug services, violence prevention & response and disability services.

Collectively, his team is on the front line for student mental health issues. They are the first ones to know whether a student is psychiatrically hospitalized, is disruptive or needs immediate financial assistance. In fact, they often refer students to the mental health counseling center. However, like the counseling center, David and his team can't meet the demand. So, many students, who don't exhibit extreme symptoms, fall through the cracks.

As I spoke to David, he described a specific bottleneck — his team couldn't share sensitive information in a timely manner. In higher education, schools are often decentralized, so each department buys their own preferred supplies and software. For example, within David's five departments, they have a student conduct software, a Title IX software, a student profile database and another student conduct software.

While this process is ok at first, it's a nightmare at scale. At MIT alone, there are over 7,000 student support appointments, 200 well-being checks, and 70 hospitalizations per year. To put those numbers into perspective, there are *only* 22 employees within David's five departments. For the most part, David and his team are fighting fires. So, it's easy to overlook a student who seemed normal.

The problem occurs when that mild case becomes an emergency. During those times, sharing information and coordinating care is essential for survival. David and his team would spend hours meeting and sifting through databases to understand what happened, why it happened and whether there were any missed warning signs.

When I heard about this problem, it was like a light bulb moment. Instead of spending hours to sift through various databases, **why not let a machine do it?**

In theory, a machine could perform the action in a fraction of the time and cost with more accuracy. More importantly, it could leverage historical data to predict future emergencies well in advance. If David and his team were to adopt this idea, they could save lives. When I proposed this idea, the administration offered to help. And so, this hypothesis was born.

3.4 Hypothesis

Disparate sources of institutional data can be used **to predict at-risk students, *even* those who don't report their mental illness.**

3.5 Limitations

The original intent was to combine data from multiple databases, or disparate sources of data, ranging from core information systems to well-being checks. By doing so, I would have created one of the most comprehensive datasets on student mental health. Under normal circumstances, many of this information may not have been accessible due to FERPA or HIPAA laws. The key challenge was to get the right data that would provide enough insights without violating the privacy of students.

However, I never had the opportunity to do so. When the COVID-19 pandemic hit, most administrators were no longer able to provide me with access. Fortunately, the Registrar's office was kind enough to honor my request. Although I don't have a complete dataset, I still can prove the hypothesis. The de-identified data, provided by the Registrar, reveals 'what' happened to the student population, but we'll have to infer 'why' it happened.

Chapter 4

Method

In order to predict at-risk students, I'm applying a standard data science methodology to help answer the following questions:

- Is the data useful?
- What test should I conduct?
- What algorithm should I use?
- How do I interpret the results?

4.1 Resources and Tools

A machine learning model is built through a combination of math and logic. I programmed the logic in Python. In conjunction, I used a number of prebuilt templates to complete the data analysis. For data manipulation, I used 'Pandas'. For scientific computing, I used 'Numpy'. For data visualization, I designed charts with 'Matplotlib' and 'Seaborn'. For machine learning, I built the model with 'Scikit-learn'. Last, I compiled and executed the code in 'Jupyter Notebook'.

4.2 Raw Data

I requested and received a de-identified spreadsheet from the MIT Registrar's Office. The 'raw data' consisted of over 15,000 students, who have graduated from MIT within the past 5 years. The spreadsheet contained more than 170,000 rows of data, which included details such as gender, major, enrollment status and grades, among others. Suffice to say, I was provided with a comprehensive dataset with relevant information. As an example, the spreadsheet looked something like the table below:

ID	Gender	Ethnicity	Major	Year	Enrollment Status	Subject	Grade	Degree
1	M	60	6	G	RE	6.862	A-	SB2006
2	F	50	2A	3	IN	2.739J	A	
3	M	88	18	2	WE	18.06	B	SB2014

Table 4.1: A visual representation of the raw, unedited spreadsheet.

4.3 Data Cleansing

Data cleansing is the process of editing the data, so that the machine learning model will work. In my process, I first made sure that all of the cells were aligned. Next, I

found missing gaps in data, which could ruin the model. Then, I translated the code words into a natural language, so that I could understand the spreadsheet:

ID	Gender	Ethnicity	Major	Year	Enrollment Status	Degree
1	Male	White American	CompSci	Graduate	Registered	Yes
2	Female	Asian American	MechE	Junior	Ineligible	No
3	Male	International	Math	Sophomore	Withdrew	Yes

Table 4.2: A visual representation of the 'cleaned', edited spreadsheet.

4.4 Defining the Objective

The purpose of the objective is to create a simple test that will either prove or disprove the hypothesis. As a reminder, the hypothesis is that ***this*** institutional data can be used to predict at-risk students. I define an at-risk student as someone who will dropout from MIT due to mental illness. However, I don't have access to mental health records. So, I have to make a few assumptions:

1. **64%** of all college dropouts left school due to mental illness [24]
2. This proportion holds true for MIT undergraduate students

Therefore,

the objective is to predict whether a student will dropout or not.

4.5 Data Refining

Now that we've defined the objective, we have to cut down the spreadsheet to fit the model. First, we'll eliminate all graduate students. Since graduate students have less structured degree paths, it may throw off the model and require a more complex algorithm. Then, we'll shrink the number of rows by taking the final semester of each student.

To further simplify, we'll remove the subject-specific information, like course grades. Course grades can produce a lot of noise, or meaningless information, for certain situations. For example, the administration may want to perform this analysis on incoming freshman. Those students have yet to earn grades, so a model that heavily weights gpa may not be as accurate.

I also converted a few columns into more useful metrics. One column determines whether a student has taken a leave of absence at any point in their academic career. I derived this formula from the enrollment statuses. The other determines whether a student dropped out or not. This column is derived from the year in which the degree was awarded.

After refining the dataset, we are left with 4,372 undergraduate students. In a spreadsheet, the data looks something like this:

ID	Gender	Ethnicity	Major	Leave of Absence	Dropout
2	Female	Asian American	MechE	Yes	No
3	Male	International	Math	No	Yes
4	Female	Hispanic American	Physics	Yes	No

Table 4.3: A visual representation of the refined spreadsheet.

4.6 Data Exploration

The goal of data exploration is to uncover patterns, characteristics or interesting details that will help build our model. We'll use charts and graphs to answer the following questions about the data:

What was the male-female ratio?

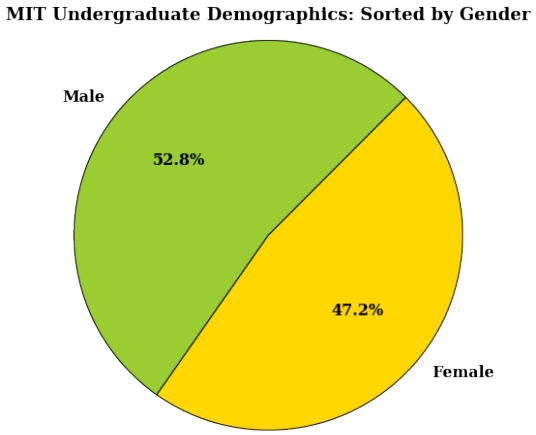


Figure 4-1: A visual representation of the gender of MIT.

How diverse was this group?

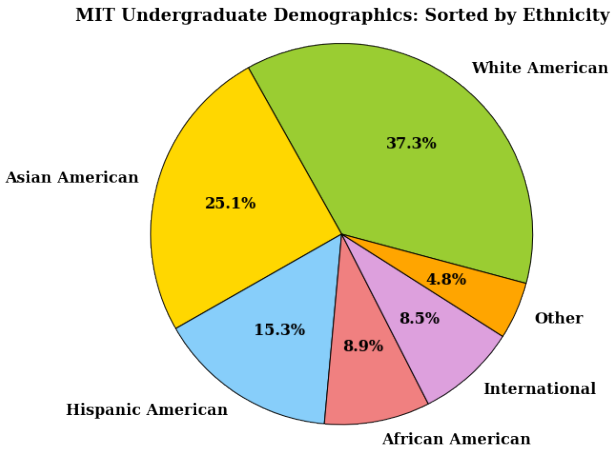


Figure 4-2: A visual representation of the ethnic diversity of MIT.

What did they study?

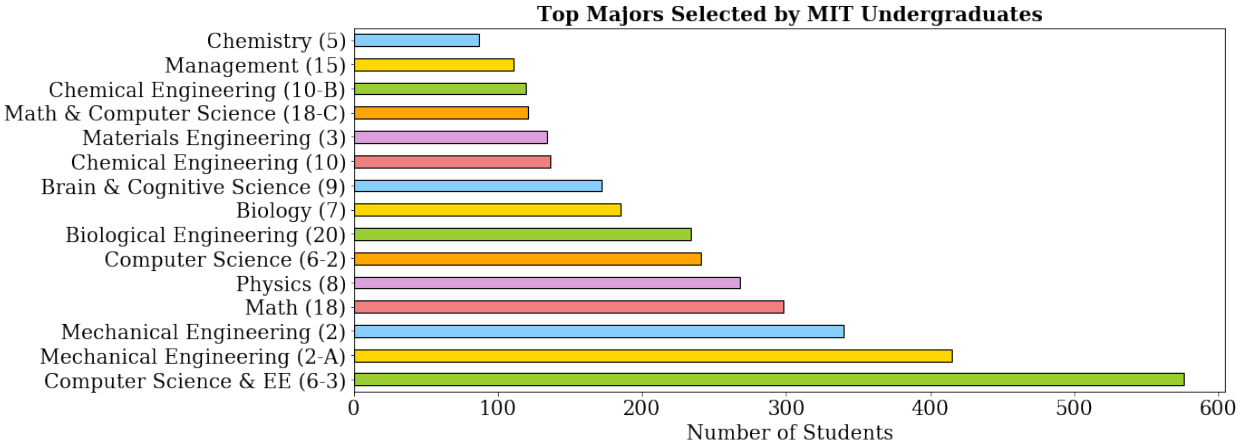


Figure 4-3: A visual representation of the majors selected at MIT.

How many students took a leave of absence?

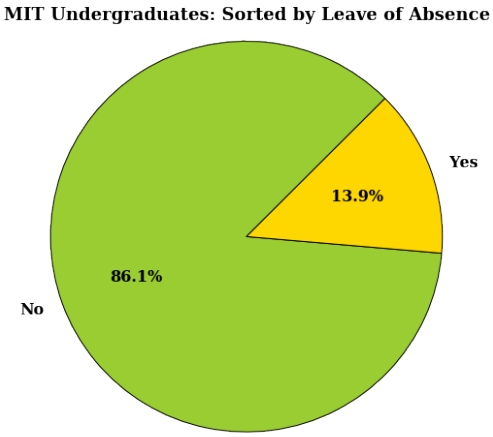


Figure 4-4: A visual representation of MIT students who took a leave of absence

How many students dropped out?

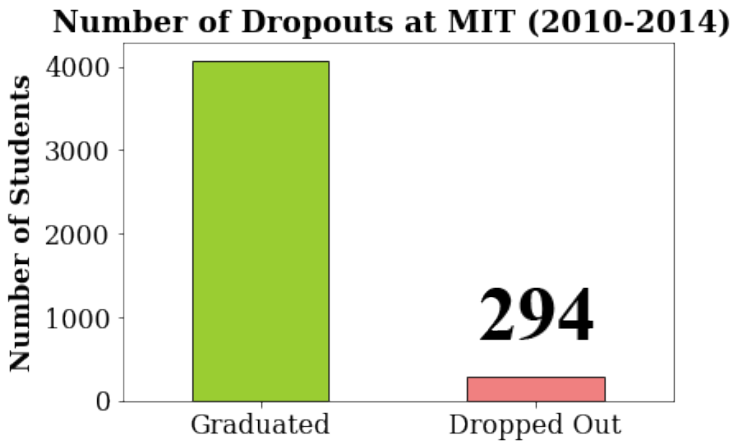


Figure 4-5: A visual representation of the number of dropouts from MIT.

How many dropouts may have suffered from mental illness?

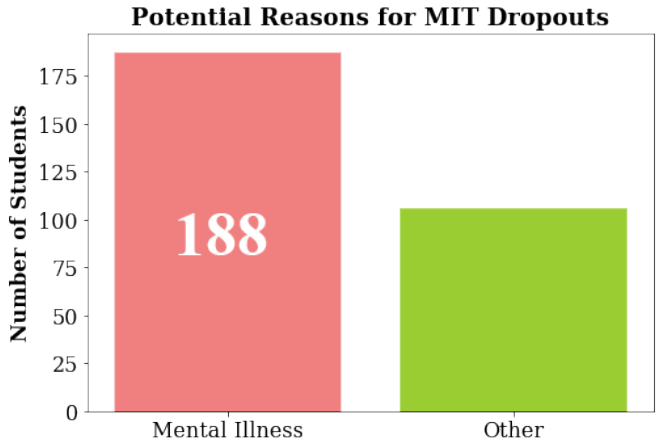


Figure 4-6: A visual representation of the potential causes of dropping out.

What was the gender ratio of MIT dropouts?

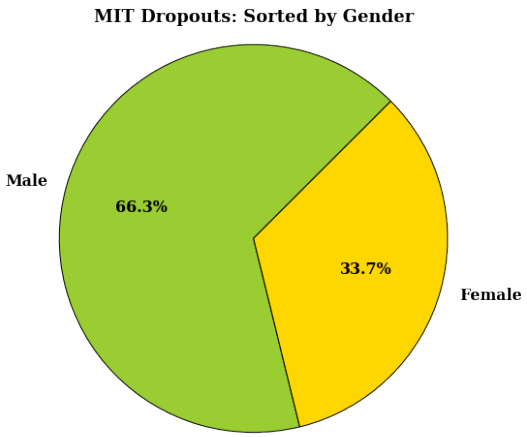


Figure 4-7: A visual representation of the gender of MIT dropouts.

What were the ethnic background of MIT dropouts?

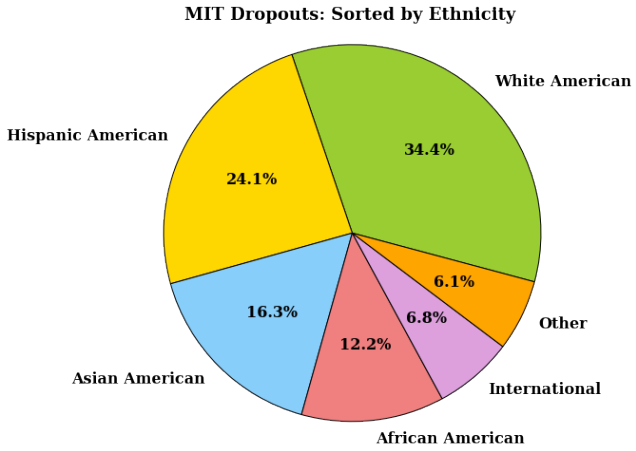


Figure 4-8: A visual representation of the ethnic background of MIT dropouts.

What did the MIT dropouts study?

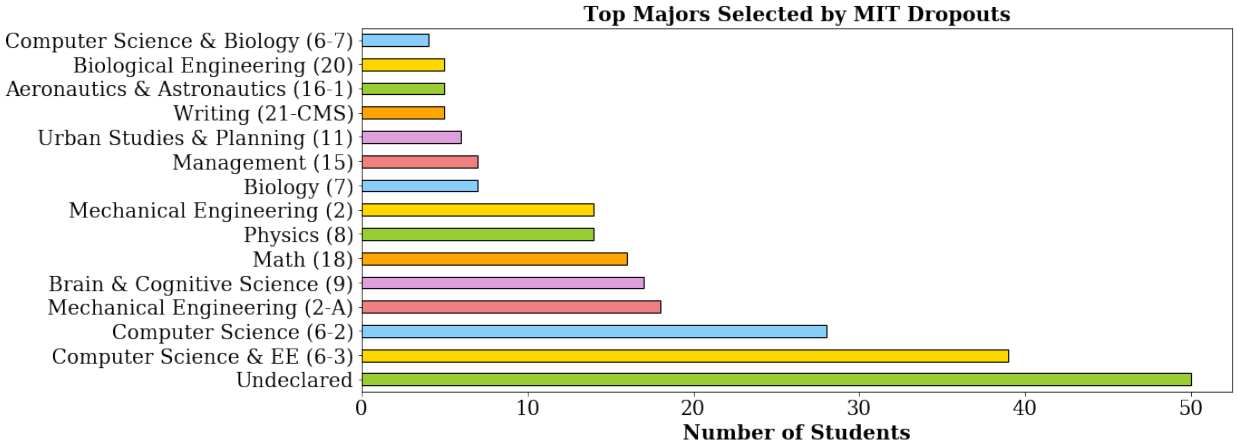


Figure 4-9: A visual representation of the majors selected by MIT dropouts.

How many dropouts took a leave of absence?

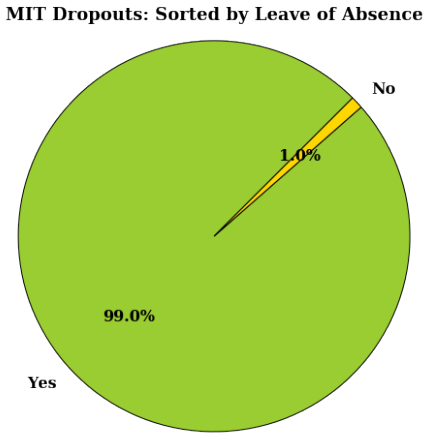


Figure 4-10: A visual representation of MIT dropouts who took a leave of absence.

Based on these data visualizations, my key takeaways are that:

1. Men are *twice* as likely as women to drop out
2. MIT's dropout rate is 6.7%, 64% of which may have left due to mental illness
3. There is a strong correlation between major and obtaining a degree
4. 10.45% of Hispanic students dropped out, the highest ratio of any ethnicity
5. **99% of students, who took a leave of absence, dropped out**

4.7 Predictive Modeling

At this stage, the objective is defined, the spreadsheet has been filtered, and we have some insight into the key drivers of the model. Now, we have to select the appropriate machine learning algorithm to test our hypothesis. Since we are testing whether a student will drop out or not, we need select a binary classification algorithm. A binary classification is the task of categorizing items into two groups. For example, yes or no, pass or fail, or fight or flight.

There are a number of techniques that can accomplish this task. However, given the time constraints, I'll be using a logistic regression algorithm. The logistic regression will use historical data to predict whether a student will drop out or not. If it works, then we'll accept the hypothesis.

The next step is to select the *'target'* column. In other words, we need to tell the model which column to predict. By doing so, the other columns will be used to predict the target column. In this case, we're selecting the 'Dropout' column. If the model predicts a 'Yes'(1) within that column, it means that the student is predicted to dropout. Conversely, a 'No'(0) means that the student is predicted to not dropout.

The last step, before building the model, is to split the data into two separate spreadsheets, one for training and one for testing. The model will learn from the training set in order to make predictions on the testing set.

Other Columns					Target Column
ID	Gender	Ethnicity	Major	Leave of Absence	Dropout
2	Female	Asian American	MechE	Yes	No
3	Male	International	Math	No	Yes
4	Female	Hispanic American	Physics	Yes	No

Table 4.4: A visual representation of the target spreadsheet.

4.8 Evaluation

In order to evaluate the logistic regression model, we'll follow a sequence of steps to determine its precision, recall and accuracy.

1) Write and execute code in Python to run the logistic regression algorithm

```
import pandas as pd
from sklearn.linear_model import LogisticRegression

Input:
    df = pd.read_csv('mit_data.csv')
    X = df['Gender','Ethnicity','Major','Leave of Absence']
    y = df['Dropout']
    regression = LogisticRegression().fit(X, y)
    predictions = reg.predict(X)
    print(predictions)

Output:
    array([0, 1, 0, ..., 0, 0, 1], dtype=uint8)
```

This code tells the regression model that the 'Dropout' column (y) is the target while the others are predictor columns (X). It then produces an output, which is a collection of 1s or 0s. As a reminder, a 1 is translated to mean 'Yes' as in 'Yes' the student will dropout of MIT.

2) Translate and export the code into a new column on the spreadsheet

Other Columns				Target Column	Machine	
ID	Gender	Ethnicity	Major	Dropout	Predictions	Translation
2	Female	Asian American	MechE	No	1	Yes
3	Male	International	Math	Yes	1	Yes
4	Female	Hispanic American	Physics	No	0	No

Table 4.5: A visual representation of the logistic regression output.

In this step, we've exported and translated the predictions into a new column inside the spreadsheet. Now, we can determine the results of the model.

3) Compare the predictions to the actual values and define the results

In order to assess the model, we have to compare each prediction with the actual value. Since there are many values, we shouldn't do this manually. Instead the machine will run a script to produce a result. However, it's important that we understand the terminology. There are four terms:

True Positive: You predicted that a student will dropout and that's true.

True Negative: You predicted that a student will not drop out and that's true.

False Positive: You predicted that a student will dropout, but that's false.

False Negative: You predicted that a student will not drop out, but that's false.

	Target Column	Machine		
ID	Dropout	Predictions	Translation	Results
2	No	1	Yes	False Positive
3	Yes	1	Yes	True Positive
4	No	0	No	True Negative

Table 4.6: A visual representation of the predicted vs actual values.

4) Build a confusion matrix

The confusion matrix, also known as an error matrix, is a specific table layout that creates a visualization of the performance of an algorithm. Essentially, we count the number of combinations and place them into a grid to assess the results.

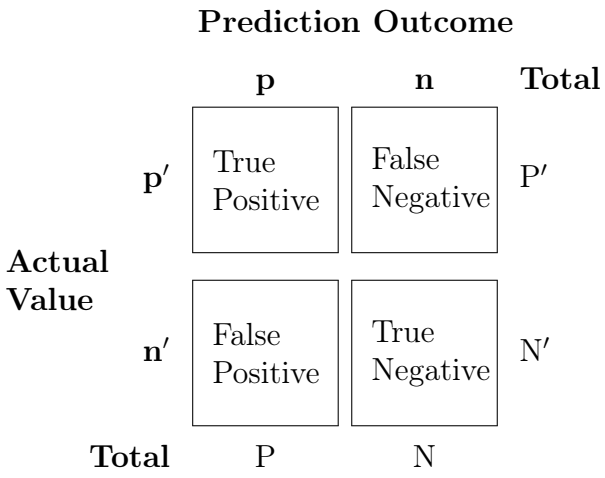


Figure 4-11: A visual representation of a confusion matrix.

5) Calculate the precision, recall and accuracy

Precision: When the model predicts a student will drop out, how often is it correct?

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

Recall: When a student drops out, how often does the model correctly predict it?

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

Accuracy: Overall, how often is the model correct?

$$Accuracy \ = \ \frac{True \ Positives + True \ Negatives}{Total}$$

Terms	Fraction	Score
Precision	1 / 2	50%
Recall	1 / 1	100%
Accuracy	2 / 3	67%

Table 4.7: A visual representation of precision, recall and accuracy.

6) Generate an ROC Curve

A receiver operating characteristic curve (ROC) is a graph that measures the performance of a binary classification system. It's derived from the confusion matrix and the terms we've defined above. Essentially, it tells us how capable the model is between distinguishes those who will drop out or not drop out. Visually, the more the curve is towards the top left, the better.

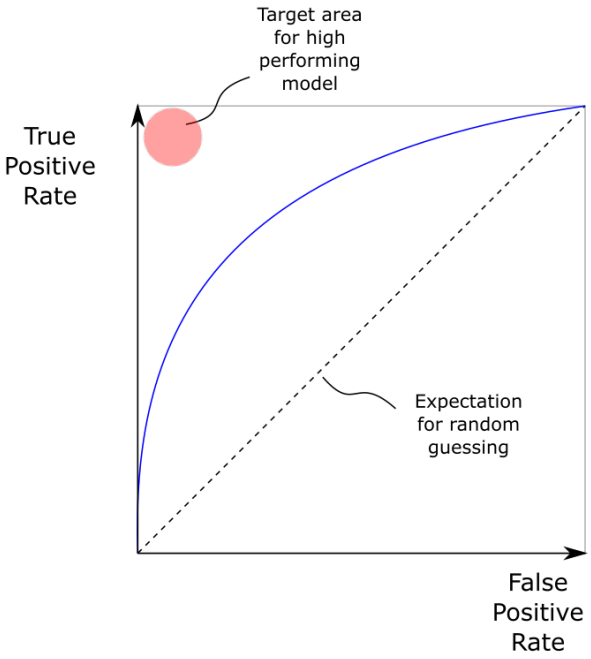


Figure 4-12: A visual representation of an ROC curve.

7) Interpret the results

After generating the predictions and creating the visual representations, we now can interpret the results. For the most part, the closer to 100%, the better. Also, the closer the curve is to the top left quadrant, the better. In certain models, we want to pay attention to particular metrics. In this case, the number of false negatives is crucial. It means that the model predicts that a student will not dropout, but they did. If the number is high, then the model is no better than a human.

Chapter 5

Discussion

5.1 Results

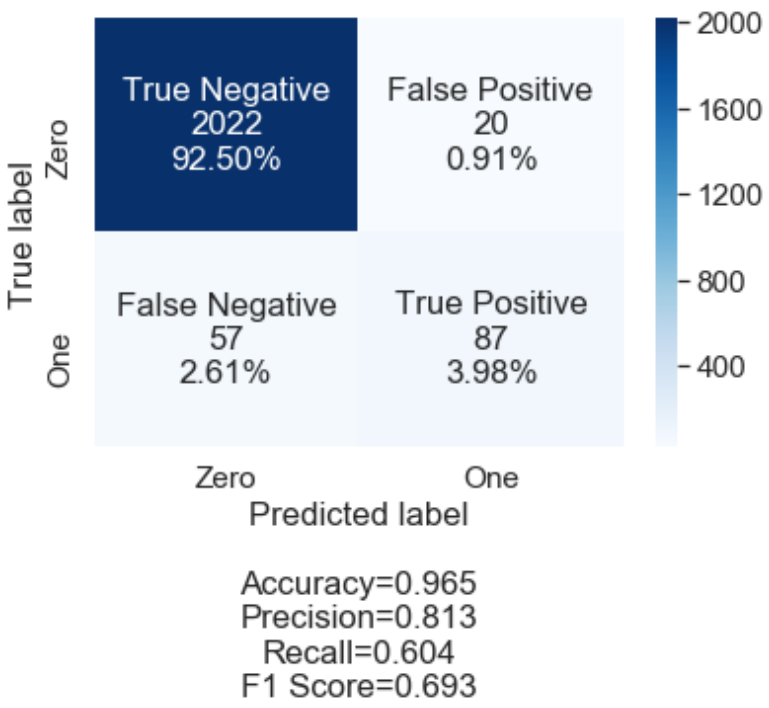
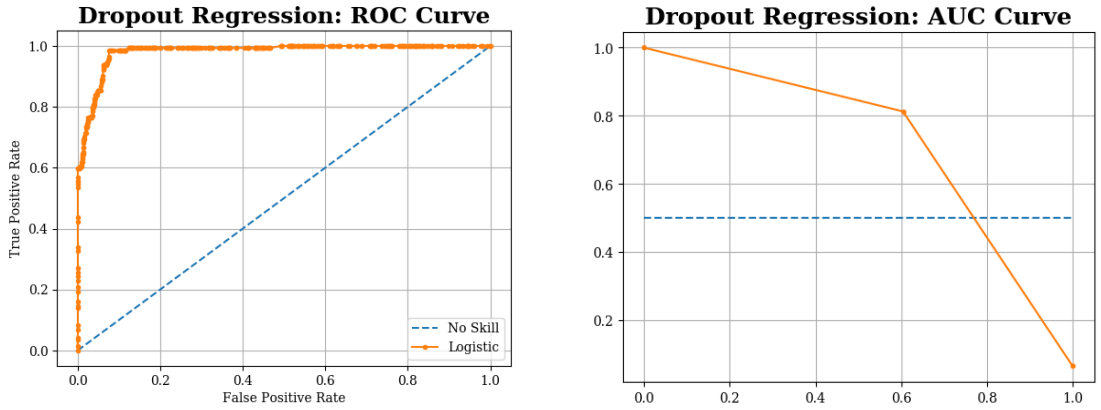


Figure 5-1: A visual representation of the confusion matrix for dropout regression.

According to the matrix above, the model was able to predict whether a student would drop out or not. The accuracy score is 96.5%, but, more importantly, the F1

score is 69.3%. The F1 Score is a combination of the precision and the recall scores of the test. This score indicates that the model is fairly accurate, but there is room for improvement.



(a) ROC Curve

(b) AUC Curve

Figure 5-2: A visual representation of the dropout regression ROC and AUC curves.

To further verify the results, the ROC curve resides in the top left quadrant of the diagram. In other words, the diagram validates the model's performance. Another graph, the AUC, or area under the curve, curve tells how much model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. By analogy, the higher the AUC, better the model is at distinguishing between dropouts and not dropouts. In this case, the AUC score is 72.2%.

In sum, we can presume that the regression model can predict the likelihood of an undergraduate student dropping out from MIT with 69.3% certainty.

Therefore, we will accept the hypothesis that disparate sources of institutional data can be used to predict at-risk students, *even* those who don't report their mental illness.

5.2 Key Insights

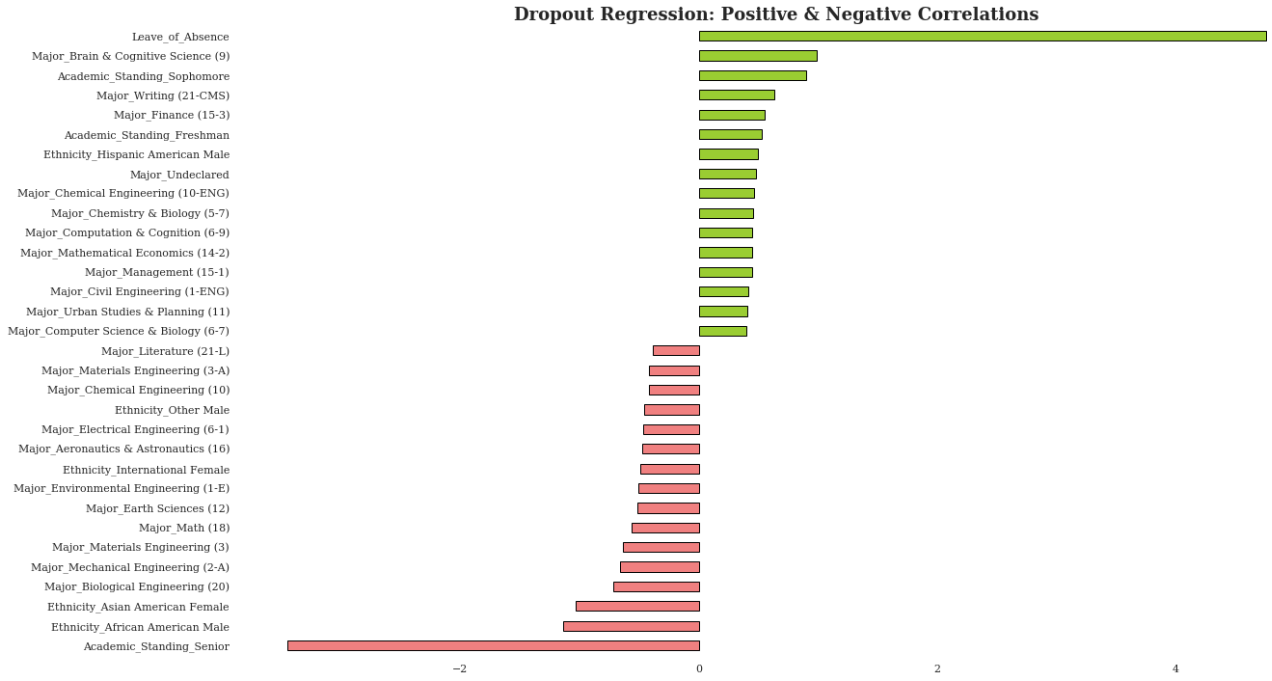


Figure 5-3: A visual representation of the dropout regression coefficients.

Based on the model and its coefficients, my key takeaways are that:

1. Taking a leave of absence is the strongest predictor of dropping out; students who take a leave of absence are almost 5 times more likely to dropout
2. Students, who majored in Brain and Cognitive Science (Course 9), were more likely to dropout than any other major
3. Hispanic American male students were more likely to dropout than any other ethnic and gender group
4. Sophomores are more likely to dropout than any other year
5. Students who did not select a major were more likely to dropout

As a reminder, the data shows us 'what' happened. In other words, it provides us with clues. In the future, we can use these insights to investigate 'why' it happened.

5.3 Conclusion

The mental health crisis has had a profound impact on higher education institutions across the nation. The rise in demand for student mental health services as well as the limited on-campus resources have led to an alarming number of leaves and dropouts.

At MIT, it's no different. In a given year, there are over 7,000 student support visits, 200 well-being checks, 100 leaves and 70 hospitalizations. Although MIT's graduation rate remains high, it does not mean that students are not suffering. As revealed in Section 3.1, there are many MIT students who fall through the cracks.

The way to combat this issue is to invest in proactive care. Unfortunately, many institutions have trouble justifying the expense. However, in Section 2.6.1, I demonstrate that institutions could earn up to $2.24 for every $1 spent.

Regardless, a viable solution for detecting at-risk students does not exist. Therefore, this thesis proposes a novel approach — leveraging machine learning to predict at-risk students from data found within the MIT Registrar's Office.

Using a binary classification algorithm, I was able to prove the hypothesis. The model was able to predict the likelihood of a student dropping out MIT with 69.3% certainty. This suggests that machine learning could be a viable and feasible solution that could predict at-risk students, *even* those who don't report their mental illness.

It is my hope that MIT and other institutions will consider adopting this technology for treating mental health. By doing so, we could prevent dropouts, leaves of absences and countless deaths due to mental illness.

5.4 Next Steps

- Refine the model — Like most datasets, there is an imbalance in classes. In other words, there are far more students who graduated than dropped out. So, the algorithm can be modified to increase the importance of certain features.

- Choose another algorithm - There are number of improvements that could be made, from including more features to changing the algorithm. A more complex model such as a neural network or principal component analysis could be also used to improve accuracy.

- Present the findings — I plan to share the details of this work with the MIT administration. Hopefully, the model will provide enough evidence to justify the inclusion of mental health data.

- Obtain mental health data — if successful, I'd like to incorporate the mental health data from student conduct software and other sources, so that I can understand 'why' students may have dropped out and if their actions were correlated to mental illness.

- Incorporate graduate students — there is information on over 11,000 graduate students that could be used to enhance the model.

- Add historical data — in order to enhance the model, I can request data from an additional number of past graduating cohorts. This will allow me to test the model on multiple generations and measure MIT's interventions over time.

- Contact other schools — while this work represents a landmark opportunity, the data within MIT may not be representative of the population at large. To improve my understanding, I'd like to test the model on other school's data.

- Build a product — the ultimate goal is to explore how to productize the model. This will require additional user research, building prototypes and user testing.

Bibliography

[1] Graduation Rate from First Institution - Digest of Education Statistics, 2018. National Center for Education Statistics (NCES) Home Page, part of the U.S. Department of Education.

[2] The Healthy Minds Study. page 25.

[3] NIMH » Any Anxiety Disorder.

[4] Physical Health and Mental health, August 2015.

[5] 2016 donorCentrics Annual Report on Higher Education Alumni Giving, 2016.

[6] 2018 Cost of Recruiting an Undergraduate Student Report, 2018.

[7] Learn About Mental Health - Mental Health - CDC, December 2018.

[8] The Silent Shortage: A White Paper Examining Supply, Demand and Recruitment Trends in Psychiatry. Technical report, Merritt Hawkins, 2018.

[9] 9 student deaths at USC since August stun campus, spark alarm, November 2019. Library Catalog: www.latimes.com Section: California.

[10] Annual collegiate mental health report examines trends and policy implications | Penn State University, January 2019. Library Catalog: news.psu.edu.

[11] First-Year Persistence and Retention for Fall 2017 Cohort. Technical report, NSC Research Center, 2019. Library Catalog: nscresearchcenter.org Section: National.

[12] Amelia M. Arria, Kimberly M. Caldeira, Kathryn B. Vincent, Emily R. Winick, Rebecca A. Baron, and Kevin E. O'Grady. Discontinuous enrollment during college: Associations with substance use and mental health. *Psychiatric services (Washington, D.C.)*, 64(2):165–172, February 2013.

[13] Victor Barr, Brian Krylowicz, David Reetz, Brian J Mistler, and Robert Rando. The Association for University and College Counseling Center Directors Annual Survey - 2011. 2011.

[14] Jill Barshay. College students predicted to fall by more than 15% after the year 2025, September 2018. Library Catalog: hechingerreport.org.

[15] Tara F. Bishop, Matthew J. Press, Salomeh Keyhani, and Harold Alan Pincus. Acceptance of insurance by psychiatrists and the implications for access to mental health care. *JAMA psychiatry*, 71(2):176–181, February 2014.

[16] Alison L. Calear and Helen Christensen. Systematic review of school-based prevention and early intervention programs for depression. *Journal of Adolescence*, 33(3):429–438, June 2010.

[17] Anthony P. Carnevale, Stephen J. Rose, and Ban Cheah. The College Payoff: Education, Occupations, Lifetime Earnings.

[18] Center for Collegiate Mental Health. 2019 Annual Report. *(Publication No. STA 20-244)*, January 2020.

[19] Colleen S. Conley, Joseph A. Durlak, Jenna B. Shapiro, Alexandra C. Kirsch, and Evan Zahniser. A Meta-Analysis of the Impact of Universal and Indicated Preventive Technology-Delivered Interventions for Higher Education Students. *Prevention Science*, 17(6):659–678, August 2016.

[20] Mona Dugo, Ben Falter, and Jamie Molnar. 2017 HECMA Membership Survey & Analysis Report, 2017.

[21] Daniel Eisenberg. The Economic Case for Mental Health Services in Higher Education. Technical Report Issue 1, The Healthy Minds Network, 2013.

[22] Daniel Eisenberg, Ezra Golberstein, and Justin B Hunt. Mental Health and Academic Success in College. *The B.E. Journal of Economic Analysis & Policy*, 9(1), September 2009.

[23] Amy C. Foerster. Involuntary Withdrawal Policies: No Room for Mental Health Stereotypes in a Fair Process, November 2019.

[24] D. Gruttadaro and D. Crudo. College students speak: A survey report on mental health, 2012.

[25] Hedden. Results from the 2011 National Survey on Drug Use and Health: Mental Health Findings,. Technical report, Substance Abuse and Mental Health Services Administration.

[26] Alketa Hysenbegasi, Steven L. Hass, and Clayton R. Rowland. The impact of depression on the academic productivity of university students. *The Journal of Mental Health Policy and Economics*, 8(3):145–151, September 2005.

[27] Heather Irias. Addiction and mental health vs. physical health: Widening disparities in network use and provider reimbursement. page 140, 2019.

[28] John Katzman. The Spending War on Student Recruitment, April 2016.

[29] Aki Kawamoto, Andres Perez Rojas, and Allison Lockard. Presenting concerns in counseling centers: The view from clinicians on the ground. *Psychological Services*, 14 (4):pp. 416–427, 2017.

[30] Bethany K. Laurence and Attorney. How Section 504 Helps Students With Physical or Mental Disabilities at School. Library Catalog: www.nolo.com.

[31] Peter LeViness, Carolyn Bershad, Kim Gorman, Lynn Braun, and Trish Murray. The Association for University and College Counseling Center Directors Annual Survey – Public Version 2018. page 73, 2018.

[32] Shannon Liao. University of Arizona tracks student ID cards to detect who might drop out, March 2018. Library Catalog: www.theverge.com.

[33] Manlu Liu and Max Cohen. CAPS Executive Director Gregory Eells died by suicide Monday morning | The Daily Pennsylvanian, September 2019.

[34] Monica Malowney, Sarah Keltz, Daniel Fischer, and J. Wesley Boyd. Availability of outpatient care from psychiatrists: a simulated-patient study in three U.S. cities. *Psychiatric Services (Washington, D.C.)*, 66(1):94–96, January 2015.

[35] Michael Maylahn. Shifting healthcare needs: Why reactive care is no longer cutting it, September 2018. Library Catalog: medcitynews.com Section: MedCity Influencers.

[36] Jenifer McKim. Harvard student's suicide prompts concern about mental health care on college campuses - The Boston Globe.

[37] Shailaja Neelakantan. 'Data Analytics Can Save Higher Education', Say Top College Bodies, November 2019. Library Catalog: edtechmagazine.com.

[38] Alison L. Neil and Helen Christensen. Australian school-based prevention and early intervention programs for anxiety and depression: a systematic review. *Medical Journal of Australia*, 186(6):305–308, 2007. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.5694/j.1326-5377.2007.tb00906.x.

[39] Paola Pedrelli, Maren Nyer, Albert Yeung, Courtney Zulauf, and Timothy Wilens. College Students: Mental Health Problems and Treatment Considerations. *Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry*, 39(5):503–511, October 2015.

[40] Daria Pelech and Tamara Hayford. Medicare Advantage And Commercial Prices For Mental Health Services. *Health Affairs*, 38(2):262–267, February 2019. Publisher: Health Affairs.

[41] Nicola Reavley and Anthony F. Jorm. Prevention and early intervention to improve mental health in higher education students: a review. *Early Intervention in Psychiatry*, 4(2):132–142, 2010. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-7893.2010.00167.x.

[42] Liz Kowalczyk Globe Staff, December 11, 2018, and 6:08 p m Email to a Friend Share on Facebook Share on TwitterPrint this Article View Comments4. Aetna settles with state over 'ghost networks' - The Boston Globe. Library Catalog: www.bostonglobe.com.

[43] Sheilynda Stewart, Doo Hun Lim, and JoHyun Kim. Factors Influencing College Persistence for First-Time Students. 38(3):9, 2015.

[44] Todd Stinebrickner and Ralph Stinebrickner. Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model. *National Bureau of Economic Research*, Working Paper 18945, April 2013.

[45] Morgan Taylor and Chessman. College Student Mental Health and Well-Being: A Survey of Presidents, August 2019. Library Catalog: www.higheredtoday.org Section: Features.

[46] Sebastian Trautmann, Jürgen Rehm, and Hans-Ulrich Wittchen. The Economic Costs of Mental Disorders. *EMBO Reports*, 17(9):1245–1249, September 2016.

[47] M. Lee. Upcraft, John N. Gardner, and Betsy O. Barefoot. *Challenging and Supporting the First-Year Student : A Handbook for Improving the First Year of College*. Jossey-Bass, 2005.