2024

Enhancing Student Graduation Rates by Mitigating Failure, Dropout, and Withdrawal in Introduction to Statistical Courses Using Statistical and Machine Learning

Shahabeddin Abbaspour Tazehkand

ENHANCING STUDENT GRADUATION RATES BY MITIGATING FAILURE, DROUPOUT, AND WITHDRAWAL IN INTRODUCTION TO STATISTICAL COURSES USING STATISTICAL AND MACHINE LEARNING

by

SHAHABEDDIN ABBASPOUR TAZEHKAND
B.Sc. Sharif University of Technology, 2012
M.Sc. Allameh Tabataba'i University, 2018
Ph.D. University of Central Florida, 2022

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Sciences
in the School of Statistics and Data Science
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Summer Term
2024

Major Professor: Morgan C. Wang

ABSTRACT

The elevated rates of failure, dropout, and withdrawal (FDW) in introductory statistics courses pose a significant barrier to students' timely graduation from college. Identifying actionable strategies to support instructors in facilitating student success by reducing FDW rates is paramount. This thesis undertakes a comprehensive approach, leveraging various machine learning algorithms to address this pressing issue. Drawing from three years of data from an introductory statistics course at one of the largest universities in the USA, this study examines the problem in depth. Numerous predictive classification models have been developed, showcasing the efficacy of machine learning techniques in this context. Actionable insights gleaned from these statistical and machine learning models have been consolidated, offering valuable guidance for instructors. Moreover, the complete analytical framework, encompassing data identification, integration, feature engineering, model development, and report generation, is meticulously outlined. By sharing this methodology, the aim is to empower researchers in the field to extend these approaches to similarly critical courses, fostering a more supportive learning environment. Ultimately, this endeavor seeks to enhance student retention and success, thereby contributing to the broader goal of promoting timely graduation from college.

This thesis is dedicated to the Woman, Life, Freedom movement in Iran
To Mahsa Amini, Kian Pirfalak, Nika Shakarami, Khodanur Lojei, Sarina Esmailzaeh, Abolfazl Adinezadeh, Hadis Najafi, Mohsen Shekari, Armita Geravand, Seyyed Mohammad Hosseini
To those who lost their lives in the movement, their families and the survivors
To Iranian schoolgirls who were poisoned in their schools and classrooms
To Iranian students who cannot access education
To Afghan female students who are being denied access to education under Islamic Theocracy
To children we lose every day in meaningless wars and conflicts
And to those children and students who don't have a voice and those who strive to give them one

ACKNOWLEDGEMENTS

I'd like to express my heartfelt gratitude to Dr. Morgan C. Wang, my advisor, who guided me through this journey. He allowed me to integrate my passion for education with the fields of Statistics and Data Science. Not only did he allow me to pursue my interests, but he also encouraged me every step of the way. I deeply appreciate all the help and guidance Dr. Wang provided, both in research and in other aspects of being a scholar. I also want to extend my thanks to Dr. Jongik Chung, from whom I learned as a student in his classes and as a graduate student receiving his feedback on my thesis committee. Dr. Rui Xie, as a committee member, also helped me understand my problem and its solution better, motivating me to improve upon my work in the future.

To my classmates and friends, I am grateful for your support. Special thanks to Md Mehedi Hasan Bhuiyan, Sandamini Seranatne, and Shahd Alnofaie, who have been my study partners and more importantly, my friends, from day one of the program. My roommate, Babak, made these past years more enjoyable and motivated me to finish my studies sooner. Milad and Sajjad, you are amazing, and I feel fortunate to have had the chance to be your friend.

To Maman Alieh and dear Hesam, thank you for your unconditional support. I love you and hope to hug you both someday soon. My dear Elham, thank you for your unwavering support throughout this journey and for pushing me forward. I cannot thank you enough.

Lastly, I am grateful for my friends from the Sharif University Basketball team: MohammadReza, Payam, Mohammad, Mohammad Ali, Ali, Ali, Behrang, AmirHasan, and Yavar. I am proud that each year this friendship grows older and stronger.

You all bring joy to my life. Thank you!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

EDM	Educational Data Mining
ML	Machine Learning
MLA	Machine Learning Algorithms
STEM	Science, Technology, Engineering, and Mathematics
XGBoost	eXtreme Gradient Boosting

CHAPTER ONE: INTRODUCTION

Introduction

Investing in education stands as a foundational pillar for national progress and individual empowerment. The virtues of educational investment transcend the mere acquisition of knowledge, fostering the comprehensive development of individuals equipped to navigate and enhance the modern world. This investment seeds the growth of a skilled workforce, nurtures innovation, and underpins social cohesion, laying the groundwork for sustainable economic prosperity and societal well-being. Recognizing the transformative power of education, governments, non-profit organizations, and international bodies allocate substantial resources towards the development and expansion of educational infrastructures. This commitment reflects a collective acknowledgment of education's role as a catalyst for positive change, driving advancements in technology, healthcare, and governance, and facilitating informed participation in democratic processes.

However, the allocation of resources towards education is not without its challenges. The finite nature of these resources necessitates a strategic approach to their distribution and utilization. It demands rigorous planning, assessment, and accountability to ensure that investments are channeled into areas where they can yield the greatest impact. This includes prioritizing access to quality education for all, enhancing teacher training, integrating technology into learning environments, and supporting educational research. Moreover, the effectiveness of educational investment is contingent upon its alignment with broader socio-economic objectives and the adaptability of educational systems to evolving global trends. As such, policy-makers and educational leaders must remain vigilant in their efforts to optimize the use of limited

resources, continuously seeking innovative solutions to meet the educational needs of diverse populations.

The evaluation and optimization of educational investments are foundational to the future of global education systems. Establishing robust mechanisms for monitoring and evaluation is not just a procedural necessity but a strategic imperative to ensure the effectiveness and relevance of educational programs. Through meticulous assessment of the outcomes of various educational initiatives, stakeholders are empowered with the data necessary to gauge program success, pinpoint areas needing enhancement, and make evidence-based decisions that shape the future landscape of education. This continuous cycle of evaluation and refinement is essential in maintaining educational programs that are not only equitable and accessible but also dynamically aligned with the evolving demands of the global workforce and societal changes. Such strategic allocation and utilization of educational resources underscore the critical role of investment in education, which serves as a cornerstone for individual development and societal advancement, promising a future marked by greater equity, opportunity, and prosperity.

Addressing the persistent challenge of program dropouts requires a multifaceted approach, recognizing the profound implications of this issue for individuals and society at large. Students who leave their educational programs prematurely face a myriad of barriers, including limited job prospects and lower earning potential, which can precipitate a cycle of economic disadvantage and increased reliance on social welfare systems. From a broader perspective, high dropout rates represent a significant impediment to national progress, stifling the development of a skilled workforce essential for economic growth and competitiveness. The societal cost of dropout rates extends beyond economics, as communities lose out on potential leaders,

innovators, and contributors who, with the right support and opportunities, could drive social change, cultural enrichment, and technological advancement. Addressing this challenge necessitates a proactive and inclusive educational strategy that identifies at-risk students early, provides targeted support and interventions, and fosters an educational environment that values diversity, inclusion, and adaptability.

In crafting solutions to reduce dropout rates and enhance the return on educational investments, stakeholders must consider the integration of comprehensive support systems, personalized learning pathways, and community engagement initiatives. Tailoring educational experiences to meet the diverse needs and circumstances of all students can create more engaging and relevant learning environments that encourage persistence and achievement. Moreover, by fostering strong partnerships between educational institutions, businesses, and community organizations, it is possible to create a cohesive support network that addresses the academic, social, and economic factors contributing to dropout rates. Through collaborative efforts, the education system can evolve to not only minimize the incidence of dropouts but also to unlock the potential of every student, thereby contributing to a more educated, skilled, and innovative society. This holistic and strategic approach to education investment and reform is imperative for building resilient, equitable, and prosperous communities worldwide, highlighting the undeniable value of education as the bedrock of societal advancement and individual fulfillment.

Addressing the intricate challenge of program and course dropouts is pivotal in the landscape of educational reform and improvement. Research underscores a significant correlation between individual course failures and the likelihood of students discontinuing their

entire academic programs. This link suggests that interventions aimed at reducing course dropout rates can serve as a critical lever in decreasing overall program dropouts, enhancing student retention, and ultimately fostering academic success. Understanding the multifaceted reasons behind course failures is the first step in a strategic approach to mitigating this issue. It demands a commitment from educational institutions to delve into the root causes of academic challenges, ranging from personal and financial hurdles to gaps in foundational knowledge or support systems.

The role of educational institutions extends beyond mere facilitators of knowledge; they must actively engage in identifying students at risk of underperformance or disengagement. Proactive identification involves leveraging data analytics and monitoring systems to flag early signs of academic distress, allowing for timely interventions. Allocating resources effectively to support these students is crucial—be it through tutoring, counseling, financial aid, or mentorship programs. Tailored support initiatives can address specific barriers to success, making the academic journey more navigable for students facing challenges. Moreover, fostering a supportive and inclusive academic environment encourages persistence, resilience, and a sense of belonging among students, which are essential factors in reducing dropout rates.

Implementing a comprehensive support system for at-risk students necessitates a collaborative effort among faculty, advisors, and support staff to create personalized intervention strategies. These strategies may include academic advising tailored to individual student needs, learning communities that provide peer support, and development workshops that enhance study skills and time management. Encouraging engagement through active learning techniques, real-world applications of course content, and opportunities for student feedback can also enhance the

learning experience, making it more relevant and engaging. By doing so, educational institutions can help students overcome obstacles, achieve success in their courses, and, crucially, maintain their trajectory toward completing their academic programs.

Ultimately, the concerted effort to reduce course and program dropout rates is not only an investment in individual students but also in the broader societal fabric. Students who successfully navigate their academic pathways contribute to a more educated, skilled, and versatile workforce, driving innovation and economic growth. Furthermore, reducing dropout rates aligns with the principles of equity and access in education, ensuring that all students, regardless of their background or challenges, have the opportunity to succeed and thrive. Educational institutions play a vital role in shaping these outcomes through their commitment to understanding, supporting, and empowering their students, laying the groundwork for a future where academic success is accessible to all.

But identifying students who are at danger is a complicated and multidimensional procedure. Analyzing a variety of variables is required, such as attendance, academic achievement, participation in class activities, and individual situations. It can take a long time and a lot of resources to complete this process, which frequently calls for significant work from administrators and instructors.

Here's where machine learning (ML) models and educational data mining (EDM) come in assistance. These tools make it possible to quickly and precisely identify students who might be in danger. EDM and ML can find patterns and insights in educational data that traditional methods might miss by analyzing large volumes of data. By using these insights, instructors can

make the best use of the limited resources available by giving focused support to kids who need it the most.

EDM and ML are particularly effective because they can analyze data from various sources, including student grades, attendance records, online learning behaviors, and even social and emotional factors. This comprehensive analysis enables a more nuanced understanding of student needs, leading to more effective interventions. Furthermore, these technologies can continuously learn and improve over time, adapting to changing educational environments and student populations.

The use of ML and EDM in teaching has enormous potential benefits. By helping educational organizations and institutions to identifying at-risk students, they may result in better student performance, lower dropout rates, and more effective use of educational resources. Moreover, by guaranteeing that assistance is given based on objective facts rather than subjective assessments, these technologies can contribute to the personalization of education. In order to overcome educational disparities and guarantee that all students, regardless of background, have the support they need to achieve, this can be especially crucial.

This study focuses on identifying learning outcomes in an introductory statistics course using machine learning models. Because of its significance in the STEM (Science, Technology, Engineering, and Mathematics) sectors, this course was selected. As the need for STEM experts and workers with STEM backgrounds grows, it is imperative that students do well in core subjects like statistics. In order to uncover variables that indicate success or dropout risk, the study will utilize machine learning (ML) models to examine student performance data. This strategy can offer insightful information about how to help students in an efficient manner,

which will ultimately lower dropout rates and increase the number of students pursuing STEM careers.

The goal of this work is to employ machine learning models to determine learning outcomes in an introductory statistics course. This course was chosen due to its importance in the STEM (Science, Technology, Engineering, and Mathematics) fields. Students must perform well in foundational courses like statistics since there is an increasing need for STEM specialists and workers with STEM expertise. The study will analyze student performance data using machine learning (ML) models to find variables that predict success or dropout risk. Insightful information about how to assist students effectively can be obtained using this technique, which will ultimately reduce dropout rates and boost the number of students choosing STEM fields.

To effectively address the issue of program and course dropouts, educational institutions must also focus on monitoring and optimizing FDW (Failure, Dropout, and Withdrawal) rates. FDW rates are crucial metrics that provide insight into the academic health and retention effectiveness within educational settings.

- **Failure Rate:** This rate represents the percentage of students who fail a course or program. It is calculated based on the proportion of students receiving a failing grade, which is determined by the institution's specific grading threshold, at the end of a term. A high failure rate can indicate challenges within the course structure, teaching methods, or student preparedness that need to be addressed to improve academic success.
- **Dropout Rate:** This rate reflects the percentage of students who discontinue a course or academic program before completing it. Students may drop out due to

various reasons, including academic difficulties, personal issues, financial constraints, or a lack of interest. Monitoring dropout rates helps institutions understand and address the factors leading students to leave their studies prematurely, which is crucial for enhancing student retention and success.

- **Withdrawal Rate:** This rate measures the percentage of students who formally withdraw from a course or program during the academic term. Unlike dropouts, withdrawals are officially documented and typically involve students informing the institution of their decision to leave. Understanding withdrawal rates helps institutions identify patterns and reasons behind students' decisions to withdraw, allowing for targeted interventions to reduce these occurrences.

For instance, consider a course with 100 enrolled students. If 10 students fail, 5 students drop out, and 8 students withdraw, the respective rates would be:

Failure Rate: 10% (10 out of 100 students failed)

Dropout Rate: 5% (5 out of 100 students dropped out)

Withdrawal Rate: 8% (8 out of 100 students withdrew)

These rates are essential for educational institutions to monitor and address. High FDW rates can indicate areas needing improvement in teaching methods, student support systems, and overall educational strategies. By focusing on these metrics, institutions can better understand where their students are struggling and implement more effective support measures.

It is worth mentioning that the Principal Investigator (PI) of the study has conducted multiple research projects at the intersection of Mathematics Education and Technology using

qualitative research methods (Abbaspour, 2022; Abbaspour & Safi, 2023; Abbaspour & Safi, 2021). As a result, he has developed a keen interest in the research question of this study.

CHAPTER TWO: LITERATURE REVIEW

XGBoost

XGBoost, an acronym for Extreme Gradient Boosting, has emerged as a groundbreaking force in the realm of machine learning, distinguishing itself through unparalleled efficiency and precision in predictive modeling tasks. This advanced algorithm, conceived by Tianqi Chen and Carlos Guestrin, has significantly enhanced the landscape of gradient boosting techniques by emphasizing optimization in both speed and overall performance. As this literature review unfolds, it will delve into the rich tapestry of XGBoost's evolution, its unique characteristics, the diverse range of applications it serves, and its competitive edge over other algorithms, while also considering the hurdles and constraints inherent to its deployment.

At its core, XGBoost is a sophisticated ensemble learning method that leverages the power of multiple decision trees to make accurate predictions. Unlike its predecessors, XGBoost incorporates several key innovations that amplify its effectiveness: a robust handling of missing data, an advanced regularization feature that mitigates overfitting, and a scalable, flexible architecture that excels in various computing environments. These enhancements not only bolster XGBoost's predictive capabilities but also contribute to its versatility, making it adaptable to a wide array of data types and analytical scenarios.

The application spectrum of XGBoost is impressively broad, encompassing fields as diverse as finance, healthcare, retail, and beyond. In finance, XGBoost has been instrumental in fraud detection and credit scoring, where its ability to handle imbalanced datasets and extract subtle patterns significantly improves decision-making processes. In healthcare, it aids in disease diagnosis and patient prognosis by accurately interpreting complex patient data. Retailers

leverage XGBoost for inventory forecasting and customer segmentation, benefiting from its nuanced understanding of consumer behavior patterns. These examples merely scratch the surface of XGBoost's potential, underscoring its transformative impact across industries.

Despite its strengths, XGBoost is not without its challenges. One notable difficulty involves hyperparameter tuning, which requires a careful balancing act to achieve optimal model performance without succumbing to overfitting or underfitting. Additionally, while XGBoost performs exceptionally well with structured tabular data, its effectiveness may diminish with unstructured data types, such as images or text, where deep learning models tend to have an advantage. These limitations necessitate a thoughtful approach to model selection and configuration, guided by the specific requirements and characteristics of the dataset at hand.

In conclusion, XGBoost stands as a testament to the ongoing evolution of machine learning technologies, offering a potent tool for data scientists and researchers seeking to unravel complex predictive modeling challenges. Its development not only marks a significant milestone in the advancement of gradient boosting methods but also sets a new standard for algorithmic efficiency and versatility. As the machine learning community continues to explore and expand upon the capabilities of XGBoost, its role in driving forward analytical excellence and innovation remains undeniably central. The journey of XGBoost, from its conceptual inception to its widespread adoption and acclaim, encapsulates the dynamic interplay of theoretical innovation and practical application that defines the cutting edge of machine learning research.

Chen and Guestrin unveiled XGBoost in their seminal 2016 paper, which not only elaborated on the algorithm's underlying mechanisms but also its design ethos aimed at addressing the scalability and efficiency challenges inherent in previous models of gradient

boosting (Chen & Guestrin, 2016). XGBoost distinguishes itself through several innovative features that significantly boost its performance. These include an efficient handling of sparse data, the incorporation of a regularized model to curb overfitting, and a scalable architecture that exploits the power of multi-core computing systems.

A cornerstone of XGBoost's methodology is its gradient boosting framework, which iteratively corrects the mistakes of previous models to refine its predictions. Unlike other models, XGBoost employs a sophisticated form of regularization, which effectively minimizes over-complexity in the model, thereby preventing overfitting—a common pitfall in machine learning (Natekin & Knoll, 2013). This approach not only enhances the predictive accuracy but also ensures the robustness of the model across diverse datasets.

The adaptability and superior predictive performance of XGBoost have seen its application across a myriad of fields, ranging from the financial industry to bioinformatics, and from computational advertising to energy forecasting. Its robustness and versatility have established XGBoost as a preferred choice for data scientists and researchers grappling with complex classification, regression, or ranking problems.

In the financial sector, XGBoost has proven instrumental in enhancing credit scoring models and fraud detection systems, allowing institutions to better predict defaults and identify fraudulent activities with remarkable precision (Zhao & Hryniewicki, 2018). The bioinformatics field has similarly benefited from the algorithm's capability to classify gene expressions accurately, thereby aiding in the identification of critical disease markers and facilitating the push towards personalized medicine approaches (Xu et al., 2018).

XGBoost's excellence in predictive modeling is not merely anecdotal but is supported by a wealth of comparative studies. These analyses consistently show XGBoost outperforming conventional models like logistic regression, support vector machines, and even other advanced tree-based methods such as Random Forests. The algorithm's supremacy is attributed to several factors, including its nuanced handling of missing data, direct processing of both categorical and numerical variables, and the implementation of advanced regularization techniques that significantly mitigate the risk of overfitting (Caruana & Niculescu-Mizil, 2006; Chen & Guestrin, 2016).

The efficiency of XGBoost is further underscored by its unparalleled scalability and parallel processing features, which ensure its swift execution even when dealing with voluminous datasets. This aspect is particularly crucial in an era where data volumes are exponentially growing, necessitating algorithms that can keep pace with the expanding scale of data.

Despite its numerous strengths, navigating the complexities of XGBoost can pose significant challenges, particularly in the realm of hyperparameter tuning. The algorithm's performance is highly sensitive to its parameter settings, and optimizing these parameters can be a daunting task that requires extensive experimentation or the application of automated optimization techniques (Bergstra et al., 2011).

Moreover, while XGBoost exhibits stellar performance on structured or tabular data, its efficacy can diminish in scenarios predominantly involving unstructured data, such as text or images. In these instances, deep learning models, which are inherently designed to capture complex patterns in unstructured data, may offer more effective solutions.

XGBoost stands as a testament to the advancements in machine learning algorithms, offering a blend of speed, accuracy, and efficiency that is unrivaled in many scenarios. Its development has significantly contributed to the predictive modeling toolkit, offering a robust solution capable of tackling a wide array of challenges across different domains. Despite the hurdles associated with parameter tuning and certain limitations in handling unstructured data, XGBoost's contributions to both the theory and practice of machine learning remain invaluable. As the field continues to evolve, the role of XGBoost in shaping future developments in predictive analytics is undoubtedly significant, underscoring its importance as a critical tool for data scientists and researchers worldwide.

The Use of XGBoost Method in Education

The deployment of XGBoost in educational environments underscores the expansive reach and transformative impact of machine learning technologies on educational outcomes and methodologies. This extensive literature review traverses the myriad ways in which XGBoost has been harnessed across the educational sector, illuminating its versatility and efficacy in various applications. From the predictive modeling of student performance to the nuanced identification of students at risk, the customization of learning trajectories, the insightful analysis of student evaluations of instruction (SEI), to evaluating the influence of role models and discerning the attributes of effective teaching, XGBoost has emerged as a pivotal tool in educational data mining. By aggregating and synthesizing research findings, this review casts a spotlight on the comprehensive and diverse applications of XGBoost within educational settings, showcasing its potential to revolutionize educational practices and interventions.

The predictive modeling capabilities of XGBoost enable educators and institutions to forecast student performance with remarkable accuracy, facilitating targeted interventions and support for students who may otherwise be overlooked. This proactive approach aids in mitigating potential dropouts and enhancing student achievement, thereby optimizing educational resources and efforts. Furthermore, XGBoost's application in the early identification of at-risk students allows for timely and personalized support strategies, fostering an educational environment where every student has the opportunity to succeed. The customization of learning pathways, another significant application of XGBoost, personalizes the educational experience, adapting learning materials and pedagogical approaches to match the unique needs and preferences of each student. This individualized approach not only enhances learning outcomes but also elevates student engagement and motivation.

Moreover, the application of XGBoost in analyzing SEI data offers valuable insights into the efficacy of instructional methods and the quality of the learning experience, providing feedback that can inform pedagogical refinements and innovations. The algorithm's ability to discern the impact of role models within educational contexts further highlights the importance of inspirational figures in motivating and guiding students toward academic and personal growth. Additionally, identifying the qualities that contribute to effective teaching through XGBoost analysis helps in recognizing and cultivating these attributes among educators, thereby elevating the overall quality of education.

The broad implications of XGBoost's integration into educational data mining extend beyond the immediate benefits to student performance and instructional quality. By leveraging the predictive power and analytical precision of XGBoost, educational institutions can embark

on a data-informed journey toward more equitable, effective, and engaging educational experiences. The insights derived from XGBoost applications can inform policy-making, curriculum design, resource allocation, and the broader discourse on educational equity and innovation. In essence, the adoption of XGBoost within educational settings exemplifies the symbiotic relationship between advanced machine learning technologies and the evolving needs of the educational sector, heralding a new era of data-driven educational excellence and reform.

The exploration of XGBoost's multifaceted applications within education reveals its significant potential to enhance and transform educational practices and outcomes. Through a comprehensive synthesis of research findings, this review underscores the algorithm's role in advancing educational data mining, contributing to a deeper understanding of the learning process, and fostering an educational landscape that is adaptive, inclusive, and forward-looking. As educational institutions continue to navigate the complexities of modern learning environments, the integration of XGBoost and similar machine learning technologies will undoubtedly play a crucial role in shaping the future of education.At the heart of XGBoost's application in education is its adeptness at handling tabular data, making it an indispensable tool for predictive modeling. Educational institutions are treasure troves of data, including demographics, engagement metrics, grades, and more, all ripe for analysis. The predictive capabilities of XGBoost allow for forecasting student success and identifying students at risk of underperforming or dropping out. Notably, Al Essa et al. (2020) leveraged XGBoost to analyze online learning behaviors, accurately predicting students' final grades. This predictive power enables educators to tailor interventions to the specific needs of students, potentially averting academic failures and dropouts.

A paramount application of XGBoost in education is the early identification of at-risk students. Through the analysis of educational data patterns and trends, XGBoost models can pinpoint students requiring additional support. This is particularly crucial in higher education, where early detection of at-risk students can significantly affect retention rates. Marbouti et al. (2016) demonstrated XGBoost's effectiveness in identifying students at risk of failing engineering courses, highlighting its capability to offer actionable insights to educators and administrators for timely interventions.

XGBoost's utility extends to creating personalized learning pathways by analyzing students' learning styles, performance history, and preferences. This dynamic adjustment of learning environments fosters a more inclusive and effective educational experience. Research has shown that XGBoost can classify students based on their likelihood to benefit from specific instructional strategies or content types, thereby optimizing educational outcomes by aligning with individual student needs.

Further extending the application of XGBoost is its use in analyzing student evaluations of instruction (SEI). Wang et al. (2009) utilized data-mining techniques, including XGBoost, to examine SEI, uncovering patterns in student feedback that inform instructional improvements and policy adjustments. Similarly, the impact of role models on student attitudes towards STEM subjects, as investigated by Evans, Whigham, & Wang (1995), underscores the significant insights predictive analytics can provide. By applying XGBoost to data from interventions such as the role model project, educators can quantify the interventions' impact, identifying key factors contributing to positive changes in student attitudes towards STEM fields.

The exploration of instructor qualities that resonate most with successful student outcomes is another domain where XGBoost provides valuable insights. Analyzing SEI data, including comments on instructor effectiveness, can help identify teaching characteristics that strongly correlate with high student achievement and satisfaction.

Despite the promising applications of XGBoost in education, challenges such as data quality, availability, and model interpretability must be addressed. The ethical use of predictive models in education also requires careful consideration to ensure equitable benefits across all student populations.

The application of XGBoost in education underscores the potential of advanced machine learning techniques to revolutionize educational practices. From enhancing predictive modeling to customizing learning pathways and improving instructional quality, XGBoost offers powerful tools for educators to enhance learning outcomes and proactively support at-risk students. The integration of XGBoost and similar technologies into educational data mining will likely play a pivotal role in shaping the future of education, necessitating careful consideration of data quality, interpretability, and ethical implications to realize the full benefits of technology across all student demographics.

The application of machine learning (ML) and data mining in the educational domain has increasingly gained prominence as institutions seek reliable methods to predict student performance and identify at-risk students. The following literature review encapsulates insights from a selection of studies, providing a comprehensive understanding of the state-of-the-art in this field.

Okereke et al. (2020) demonstrated the use of Decision Trees, a Machine Learning Algorithm (MLA), to parse through data from 103 first-year Computer Science students at the University of Nigeria, Nsukka. They underscored the pivotal role of feature selection, accomplished via Rapid Miner, in managing the high number of predicting variables. Their findings stress that accuracy in predicting student performance is heavily dependent on the nature of the datasets rather than the Machine Learning Algorithms (MLA) employed.

Parallel to this, Bhutto et al. (2020) explored the use of supervised machine learning algorithms, like the support vector machine and logistic regression, to predict student performance. They posited that ML could assist educational institutions in classifying students' performance for targeted interventions. Their results also pointed to the Sequential Minimal Optimization algorithm's effectiveness over logistic regression, reinforcing the importance of selecting optimal algorithms and features, including teacher's performance and student's motivation, for the predictive model.

Alyahyan and Düştegör (2020) discussed machine learning techniques as vital tools for prediction, advocating for their utility in improving student success in higher education. They provided educators with systematic guidelines on applying data mining methods, including decisions ranging from defining student success to selecting appropriate ML methods. This study serves as an indispensable reference for educators aiming to harness data mining in predicting student success.

In a similar vein, Lykourentzou et al. (2009) presented an e-learning dropout prediction method employing ML techniques such as feed-forward neural networks, support vector machines, and probabilistic ensemble simplified fuzzy ARTMAP. They emphasized the

advantage of using electronic data from e-learning platforms for dynamic and adaptable predictions of at-risk students, thus fostering proactive interventions.

The role of kernel methods in educational databases was also emphasized, as they proved adept at processing large, highly dimensional, non-linear, and non-separable data. Studies revealed positive correlations between the application of such techniques and prediction outcomes, underscoring the significance of kernel methods in education data mining.

Lastly, in the quest for understanding and improving predictive models, the research has repeatedly highlighted the impact of feature selection on model efficacy. Alyahyan and Düştegör (2020) delineated how feature selection facilitates reduced computation time, improves prediction performance, and enables a better understanding of the data, effectively contributing to the advancement of educational data mining and its practical applications.

In conclusion, the synthesis of these studies demonstrates a clear consensus on the utility of ML and data mining in predicting student performance. The importance of dataset quality, feature selection, and the thoughtful application of specific ML algorithms emerges as critical for the development of accurate predictive models. These collective insights contribute significantly to the field, supporting educators and institutions in their efforts to identify at-risk students and thus mitigate potential educational failures.

CHAPTER THREE: DATA AND METHODOLOGY

In the realm of predictive modeling, feature engineering plays a pivotal role in enhancing the predictive power of models by extracting meaningful insights from raw data. This section delves into the process of feature engineering employed in this study to augment the dataset collected from introductory statistics courses. The PI investigates the structure and characteristics of data collected from introductory statistics courses offered at an institution in the southeastern United States over a three-year period (2021-2023). The primary aim is to analyze the data's organization, its variables, and its suitability for building predictive models. Data collection involved reaching out to the Institutional Knowledge Management department, obtaining grade books and institutional data, and then processing them for analysis. The study focuses on exploring the variability in data across different course offerings and semesters, identifying patterns, and assessing the data's adequacy for predictive modeling. The findings provide insights into the complexities of managing and structuring educational data for analytical purposes.

In today's data-driven world, understanding the structure and characteristics of data is paramount for effective decision-making and predictive modeling. This thesis delves into the data collected from introductory statistics courses offered at a southeastern US institution, aiming to elucidate its structure and suitability for predictive analytics. The introductory statistics course serves as an ideal setting for studying educational data due to its standardized curriculum and widespread enrollment.

The data collection process involved collaboration with the Institutional Knowledge Management (IKM) department to procure grade books and institutional data pertaining to

introductory statistics courses offered between 2021 and 2023. The rationale behind selecting this timeframe was to utilize the first two years for model training and the third year for testing. The collected data encompassed various parameters such as quizzes, homework assignments, exams, attendance records, final grades, and scores. Notably, the data exhibited variability in terms of available features and course offerings, with some courses emphasizing certain assessments over others.

Upon collection, the raw data underwent preprocessing to ensure consistency and reliability. This involved standardizing data formats, handling missing values, and filtering out irrelevant groups, particularly those with fewer than 30 students. The data that the PI received was organized into 42 distinct CSV files, representing 54 different groups of the introductory statistics course across the three-year period. Each file contained a comprehensive array of variables reflecting students' performance and demographic information.

The exploratory data analysis (EDA) phase aimed to elucidate the inherent structure and patterns within the dataset. Descriptive statistics, visualizations, and correlation analyses were employed to gain insights into the distribution of variables, identify outliers, and discern any discernible trends or associations. Moreover, comparisons were made across semesters and course offerings to assess the variability and consistency of data.

An integral aspect of this thesis is evaluating the data's structure and its suitability for building predictive models. This entails assessing the predictive power of various features, determining the most relevant variables, and identifying potential challenges or limitations. Furthermore, considerations such as model interpretability, generalizability, and scalability are crucial in selecting an appropriate analytical approach.

By purposefully selecting the target population and features of them that was accessible the PI tried to provide a comprehensive analysis of the structure and characteristics of data collected from introductory statistics courses at a southeastern US institution. By elucidating the nuances and complexities inherent in educational data, this study offers valuable insights for educators, policymakers, and data analysts alike. Moving forward, further research could explore advanced modeling techniques, longitudinal analyses, and the integration of additional data sources to enhance predictive accuracy and inform evidence-based decision-making in education.

This thesis serves as a foundational exploration into the structure and characteristics of educational data, particularly within the context of introductory statistics courses. By employing rigorous methodologies and analytical techniques, it aims to contribute to the broader discourse on data-driven decision-making in education.

Data Structure

The bar graph presented illustrates the distribution of students passing and failing a specific course. It reveals a significant discrepancy between the number of students who passed versus those who failed. Specifically, a total of 10,750 students successfully passed the course, as indicated by the towering purple bar. In stark contrast, only 1,075 students failed the course, represented by the much shorter teal bar. This data clearly suggests a high success rate for the course, with approximately 90.9% of the students achieving a pass. The substantial difference in the magnitude of the two bars visually emphasizes the course's effectiveness in facilitating student success. This analysis could be used to infer several aspects of the course design, including the adequacy of instructional methods and materials, or perhaps the leniency of

grading systems. Further investigation might be required to dissect the underlying causes of this distribution, which could include reviewing teaching methodologies, assessment strategies, or even pre-course preparation and student selection processes.

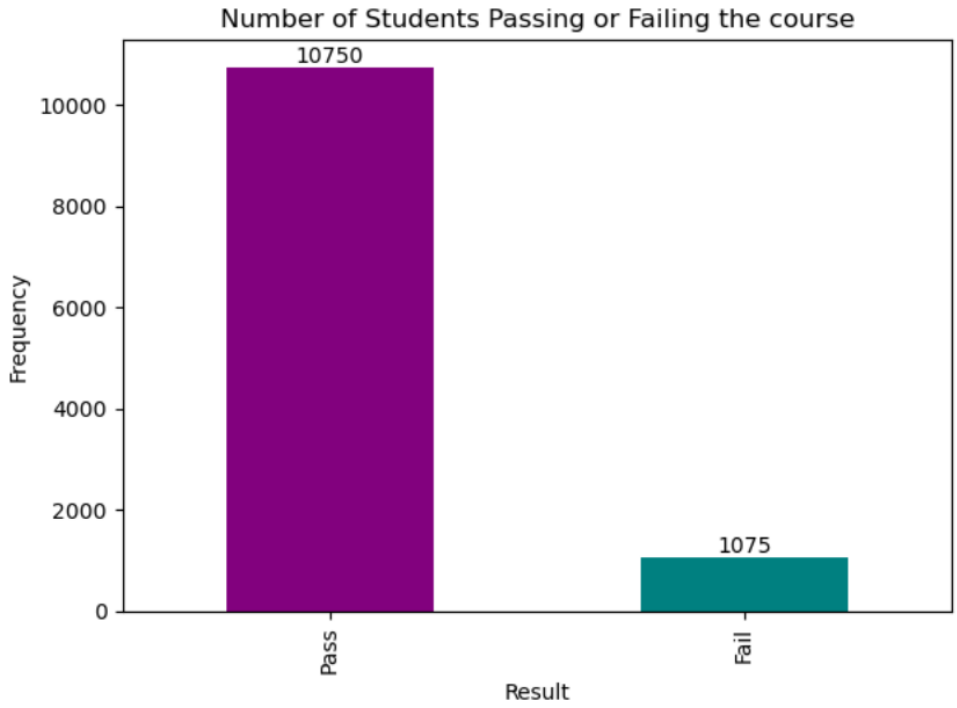

Figure 1. The number of students who pass or fail the course

The following provided bar graph represents the gender distribution of students within a specific academic program or institution, showing a distinct difference in enrollment numbers between female and male students. The graph indicates that there are 6,684 female students compared to 5,138 male students. This visualization effectively highlights a gender imbalance, with female students outnumbering their male counterparts by a significant margin.

This disparity invites a deeper exploration into the factors influencing gender distribution within the academic environment. Possible areas of inquiry might include the nature of the program's outreach and recruitment efforts, the perceived gender alignment with the field of

study, and existing support structures that may favor one gender over another. Additionally, this data can prompt discussions regarding the inclusivity of the educational setting and whether adjustments are necessary to foster a more balanced gender ratio.

Analyzing such data is crucial for educational policymakers and administrators as it provides insights that can help in developing strategies to encourage more equitable participation across genders. This might include targeted recruitment efforts, scholarship programs tailored to underrepresented genders in certain fields, or revisions in marketing strategies to appeal more broadly to prospective students. The goal would be to ensure that all students have equal opportunities and motivations to enroll and succeed in their chosen fields of study.

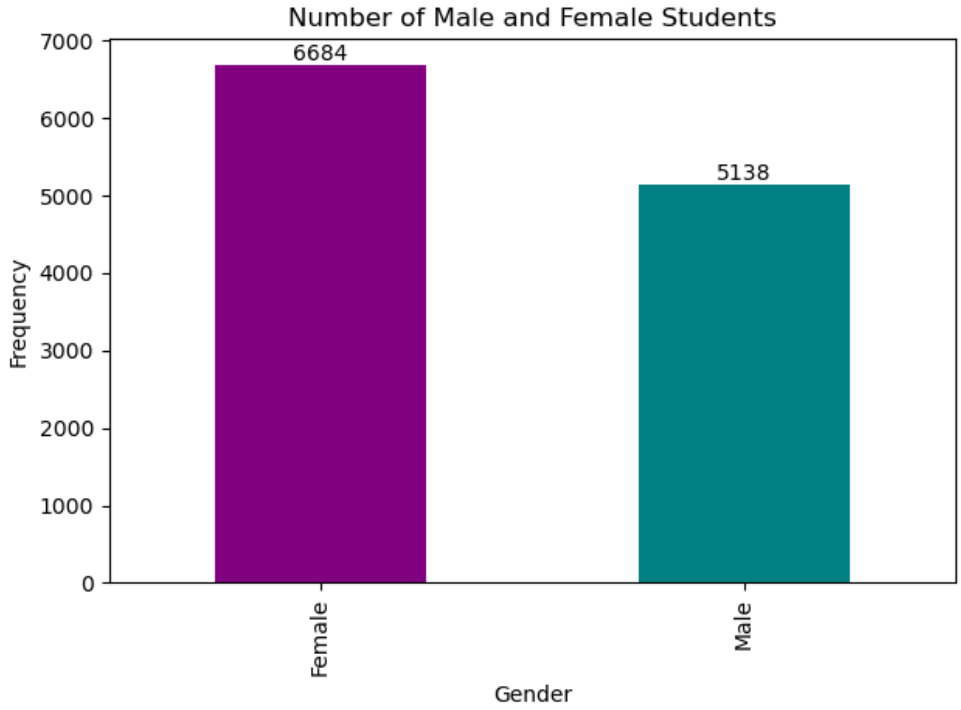

Figure 2. The number of male and female

In the following figure, the bar graph displayed illustrates the pass and fail counts by gender for a specific academic course or program. This visual representation indicates that a total

of 6,684 female students and 5,138 male students were enrolled, with the majority in each gender category successfully passing the course. Specifically, 6,076 female students and 4,672 male students passed, corresponding to pass rates of approximately 90.9% and 90.9%, respectively. The fail counts stand at 608 for females and 466 for males, representing about 9.1% of each gender group.

This distribution offers valuable insights into the comparative academic performance and success rates between genders within the course. The similarity in pass rates suggests that both genders perform relatively equally in terms of achieving course outcomes, despite the higher enrollment numbers for females. The graph effectively conveys the success of both genders, underscoring that gender does not appear to be a differentiating factor in course outcomes in this context.

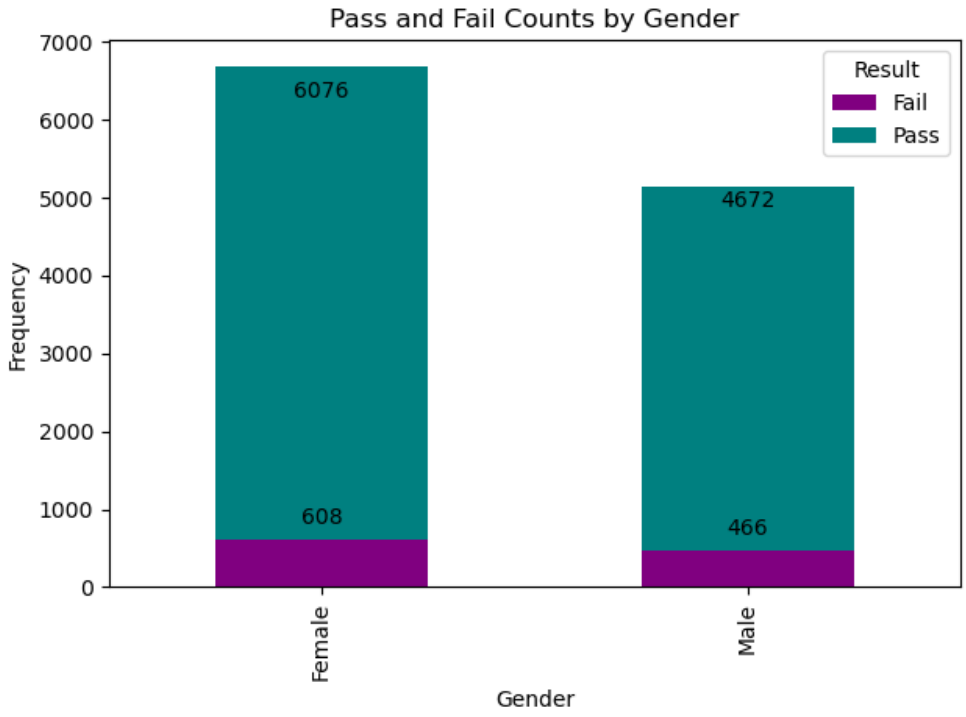

Figure 3. Pass and Fail counts by gender

Pass and fail percentages for three distinct academic semesters—Fall, Spring, and Summer—are illustrated through the distribution shown in the chart. While the graph combines data from both STEM and non-STEM fields, it does not visually differentiate between the two. Each column on the chart corresponds to a semester and demonstrates a high pass rate, with a smaller section at the bottom for those who did not succeed.

The consistent achievement across the various terms, including the typically shorter and more challenging Summer semester, suggests that the educational support systems and program structures are effectively catering to a range of learning styles and speeds.

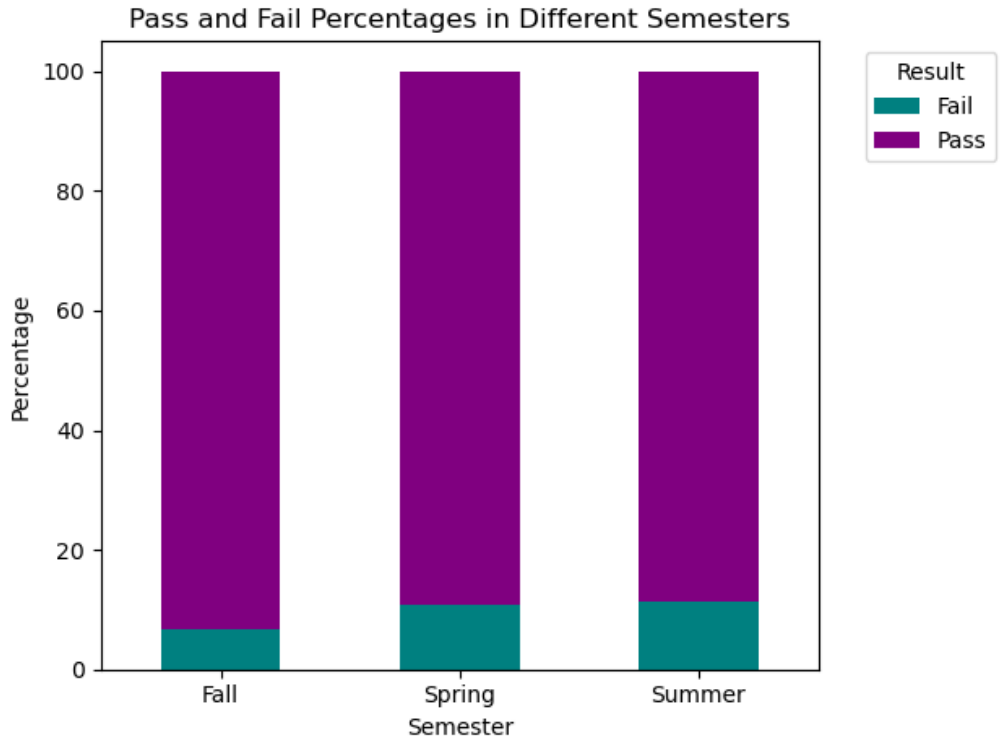

Figure 4. Pass and Fail percentages in different semesters

The chart displays the pass and fail percentages for students over three consecutive years: 2021, 2022, and 2023. Each column corresponds to one of these years, showing a high

percentage of students who successfully passed, indicated by the prominent purple sections, while the smaller teal sections at the base represent those who failed.

Remarkably, the success rates remain consistently high from 2021 through 2023, suggesting that the educational quality and support mechanisms are stable and effective across these years. The consistent pass rates across different academic years could be indicative of effective teaching methodologies, comprehensive student support services, and a curriculum that aligns well with student abilities and learning objectives. This stability in educational outcomes is crucial for maintaining confidence in the academic programs offered, indicating a robust educational environment that fosters student success year after year.

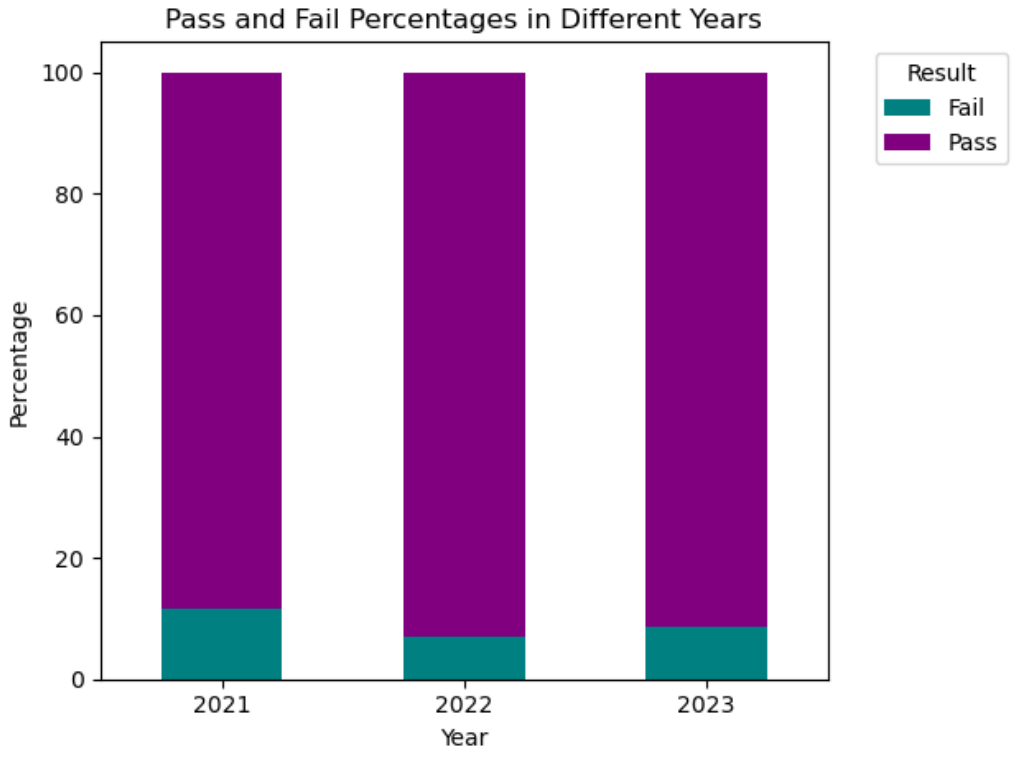

Figure 5. Pass and Fail percentages in different years

Ideas Behind Data Manipulation, Wrangling and Cleaning

In collaboration with the advisor, the primary investigator (PI) decided on defining new features to enrich the dataset's predictive capabilities. Given the variability in assignment types and instructors' teaching methodologies, it was imperative to devise features that encapsulated students' performance trends and patterns across different assessment categories. For instance, for each category of assignments (homeworks, quizzes, exams, etc.), three new features were defined:

- Assignment Trend: Reflecting students' performance trajectory throughout the semester.
- Minimum Score: Capturing the lowest score attained by students on assignments.
- Maximum Score: Indicating the highest score achieved by students on assignments.

These features aimed to provide a holistic understanding of students' engagement and proficiency levels across various assessment types.

To facilitate meaningful comparisons and analyses across different assignments and courses, all grades were standardized to a common scale ranging from 0 to 100%. This standardization process ensured uniformity in grading metrics, thereby enabling fair comparisons and aggregations across disparate datasets. Moreover, standardizing grades eliminated biases arising from differences in grading methodologies adopted by instructors, promoting consistency and reliability in the dataset.

Instructors' variations in assigning weights to different assignment categories necessitated the incorporation of assignment weights into the dataset. Notably, each instructor designed courses with distinct weighting schemes tailored to their pedagogical preferences and course objectives. Consequently, the dataset was augmented with assignment weightings corresponding

to each assignment category, thereby capturing the relative importance assigned to different assessments by instructors.

Furthermore, the dataset accounted for instructor-specific variations in assignment weights, ensuring granularity and fidelity in modeling students' performance. Instructors' idiosyncratic approaches to course design, including adjustments in assignment weights across semesters or academic years, were meticulously documented and incorporated into the dataset. This meticulous approach enabled the modeling of nuanced variations in students' performance attributable to instructor-specific factors.

In conclusion, feature engineering and standardization play a pivotal role in enriching the dataset's predictive capabilities and mitigating biases arising from instructor-specific variations in grading methodologies. By defining new features, standardizing grades, and incorporating assignment weights, the dataset becomes more robust and conducive to accurate predictive modeling. Moving forward, continued refinement and augmentation of features will bolster the dataset's utility and empower educators and policymakers with actionable insights derived from educational data analytics.

Data Preparation

The foundation of a robust data analysis in educational research lies in meticulous data preparation. This chapter elucidates the comprehensive data wrangling and cleaning process employed to transform raw, unstructured academic performance data into a clean, structured format ready for comprehensive analysis. Through this rigorous process, the dataset, encompassing homework, quizzes, exams, extra credits, and demographic information, was refined for in-depth analysis aimed at understanding and enhancing educational outcomes.

Data Collection and Initial Assessment

The initial stage involved aggregating a diverse set of student performance metrics, including but not limited to, homework scores, quizzes, exams, and demographic data. This aggregation process required meticulous attention to detail to ensure data completeness and accuracy, setting the stage for the subsequent cleaning and transformation processes:

- Source Identification: Data was sourced from various educational platforms and tools, each contributing unique metrics relevant to student performance and engagement.
- Preliminary Data Audit: A thorough audit was conducted to assess the data's initial quality, identifying issues such as missing values, inconsistencies, and outliers that could potentially skew analysis results.

The data cleaning process was both broad and detailed, addressing several critical aspects to enhance data quality:

- Standardization of Column Names: Renaming columns to intuitive, standardized titles facilitated easier navigation and manipulation. This process involved converting cryptic identifiers into clear, descriptive labels reflecting the nature of the data they contained.
- Removal of Redundant Data: Unnecessary columns, including those not relevant to the study's objectives or containing redundant information, were systematically identified and removed. This step was crucial for focusing the analysis on significant variables and streamlining the dataset.
- Missing Data Management: Different strategies were employed to handle missing data, chosen based on the nature of the missingness and the potential impact on analysis. Strategies ranged from imputation, using statistical methods to estimate missing values, to complete case analysis, excluding records with missing data.

Normalization and scaling were critical in standardizing the scores across different assessments, enabling a uniform evaluation framework. This step adjusted raw scores to a percentage scale, facilitating direct comparisons across various types of assessments and ensuring comparability in evaluation metrics.

To extract more nuanced insights from the data, new variables were engineered:

- Trend Analysis: Patterns of performance over time were identified through calculating trends across sequential assessments for each student. This approach highlighted improvements, consistencies, or declines in scores, providing a dynamic view of performance.
- Statistical Feature Creation: Calculating minimum, maximum, and average scores for different assessment categories offered a snapshot of the range and central tendency of student performance. These features were invaluable for identifying outliers and understanding distribution characteristics.

The dataset's concluding phase involved further refinement:

- Column Reassessment: After the initial round of feature engineering, some columns were reassessed. This likely involved the removal or adjustment of features that did not contribute meaningful insights, ensuring the final dataset was focused and relevant.
- Custom Ordering of Columns: To enhance the dataset's logical flow and readability, columns were reordered. This facilitated understanding and analysis, ensuring similar types of data were grouped together.
- Exporting the Cleaned Dataset: Marking the transition from data preparation to analysis, the final cleaned dataset was exported in a format suitable for analysis.

This chapter has outlined the methodical process of data wrangling and cleaning employed to prepare a comprehensive dataset for analysis in educational research. By ensuring data integrity, relevance, and coherence, this process lays a robust foundation for the subsequent analytical or modeling tasks, embodying best practices in handling, cleaning, and preparing complex datasets for insightful analysis.

This chapter outlines the comprehensive data preparation and cleaning process undertaken across various reports analyzing student performance metrics. The datasets analyzed encompass a wide range of academic performance indicators, including homework, quizzes, exams, and demographic information, requiring meticulous transformation to facilitate detailed analysis. The process involved several key steps: initial data loading and cleanup, normalization and standardization, advanced feature engineering, and final dataset adjustments, each contributing significantly to enhancing the dataset's analytical utility.

The data preparation process began with the initial loading of datasets into a pandas DataFrame, a crucial step that set the stage for all subsequent data manipulation and cleaning actions. This step involved examining the structure and scale of the datasets, followed by systematic column renaming for clarity, dropping of redundant columns not contributing to the core analysis objectives, and addressing any missing values. These actions were vital for streamlining the datasets, focusing the analysis on impactful variables, and ensuring ease of navigation and manipulation within the dataset.

Normalization and Standardization

A critical aspect of the data preparation was the normalization of scores across different assessments to a 100-point scale, ensuring uniform evaluation. This involved detailed calculations to adjust raw scores based on predetermined divisors, allowing for direct

comparability across various types of assessments. The normalization process set a standard framework for evaluation, ensuring that all performance metrics were assessed on a uniform scale.

To deepen the analysis, the reports implemented several advanced feature engineering steps. This included trend analysis using linear regression to calculate trends across homework and lecture review scores, creation of statistical features such as minimum, maximum, and trend metrics for each category of assessment, and custom calculations that demonstrated innovative approaches to capturing student performance trends. These steps enriched the dataset with nuanced insights into student performance dynamics, showcasing a meticulous approach to understanding the dataset.

Methodology

This study employs a multi-faceted approach to predictive analytics in educational settings, utilizing machine learning models—XGBoost and Decision Tree—to predict academic performance from data collected from introductory statistics courses between 2021 and 2023. Categorical variables were encoded numerically, transforming the dataset into a machine-readable format for regression and classification analyses. Four models were configured and evaluated for this purpose.

XGBoost Classifier

- An XGBoost Classifier was employed for its robust performance in classification tasks, with hyperparameters set to prevent overfitting and ensure efficient learning.
- The model was trained on data from 2021 and 2022, with 2023 reserved for testing.
- The classifier was assessed on metrics such as accuracy, precision, and F1 score to gauge its performance in correctly classifying students' final grades.

Decision Tree Classifier

- A Decision Tree Classifier was used as a baseline for its interpretability and ease of visualization. It was trained to categorize students' final grades based on learning patterns.
- The classifier was evaluated for its accuracy, precision, and F1 score, providing insights into its categorical prediction abilities.

XGBoost Regressor

- The study utilized an XGBoost Regressor, known for its efficacy in regression tasks, to forecast continuous outcomes based on historical data.
- The model's performance was quantified using MSE, RMSE, and R^2 scores, reflecting its capability to minimize errors and explain variance.

Model Performance

Model performance was evaluated using both regression metrics (MSE, RMSE, and R^2) and classification metrics (accuracy, precision, and F1 score), offering a comprehensive assessment of the models' prediction abilities.

The application of both classification and regression models allowed for a holistic analysis of the educational data. XGBoost showcased strength in both domains, while the Decision Tree offered an interpretable model that could easily classify grades. Insights gained from this study underline the importance of selecting the right model based on the analytical goal—whether it be predicting continuous outcomes or classifying categorical variables.

CHAPTER FOUR: DATA ANALYSIS

This chapter delves into the analytical methodologies and comparative evaluations of statistical models designed to forecast student performance based on educational data. Utilizing a dataset comprising several years of student academic records, including coursework scores and demographic information, this chapter explores the application of three distinct predictive models: Decision Tree Classifier, XGBoost Classifier, and a custom implementation of Logistic Regression through gradient descent. The efficacy of each model is scrutinized through a series of metrics to establish their predictive accuracy and suitability for real-world educational settings. This analysis not only highlights the technical aspects of the models but also considers their practical implications in educational systems, aiming to provide a toolset for educators to improve student outcomes based on predictive insights.

The initial dataset encompassed a range of features such as students' homework scores, quiz and exam trends, and final grades over multiple academic terms. To prepare this data for analysis, extensive preprocessing steps were undertaken. These included cleaning data to remove inconsistencies, encoding categorical variables to numerical formats using methods like one-hot encoding, and handling missing values to ensure the integrity of the analyses. This preprocessing not only improved the quality of data but also enhanced the reliability of the models' outcomes. Specific algorithms were utilized to detect and impute missing values, while feature selection was carefully conducted to include variables most predictive of student success, such as participation in class activities, frequency of homework submission, and historical academic performance.

Overview of Models

The Decision Tree Regressor

The Decision Tree Regressor is a fundamental yet powerful machine learning model widely used for its simplicity and interpretability. This model operates on the principle of recursive partitioning, breaking down a dataset into smaller and smaller subsets while simultaneously developing an associated decision tree. The final result is a tree with decision nodes and leaf nodes that represent predictions based on the input features.

In this study, the Decision Tree Regressor was applied to predict the 'Final Score' of students, which quantifies their academic performance on a scale from 0 to 100. This target variable is classified as a ratio variable because it has a meaningful zero (indicating no mastery of the course material), consistent intervals, and the ability to express one score as a multiple of another, which is essential for the model's regression tasks.

Decision Trees model the decision-making process by learning simple decision rules inferred from the training data features. The structure of a Decision Tree is straightforward: it splits the source set into subsets based on an attribute value test. This process is repeated recursively on each derived subset in a manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target variable, or when splitting no longer adds value to the predictions. The model’s simplicity can be a significant advantage when explaining the outcome to a lay audience.

The Decision Tree Regressor, with its intrinsic simplicity and direct interpretability, serves as an effective tool for predictive modeling in educational settings. While it offers substantial insights into the factors influencing student performance, its utility might be limited by its tendency to overfit, especially with more complex datasets. Future work could explore tree

pruning strategies or ensemble methods like Random Forests to overcome these limitations and enhance model robustness.

Table 1. The Decision Tree Regressor evaluation metrics

Metric	Value
MSE	**116.78**
RMSE	**10.81**
R^2	**0.520**

The XGBoost Regressor

The XGBoost Regressor stands out as a premier choice in the domain of ensemble learning, specifically within the gradient boosting framework. XGBoost, which stands for eXtreme Gradient Boosting, utilizes both tree-based learning architectures and linear models to deliver powerful predictive insights. Renowned for its performance and efficiency in various predictive modeling competitions, XGBoost is particularly favored for its robust handling of diverse datasets, from small to extremely large scales.

For this analysis, the XGBoost model was tasked with predicting the 'Final Score' of students, a continuous variable representing the percentage of marks obtained. This target variable is treated as a ratio variable, which not only defines a zero point (indicating no knowledge or mastery of the evaluated content) but also allows for meaningful and absolute comparisons between different measurements. This scale is crucial for quantitative regression tasks, enabling precise educational assessments.

XGBoost improves upon traditional gradient boosting methods by incorporating several advanced features that enhance model training and generalization. These features include built-in regularization—which helps in reducing overfitting—high flexibility in tuning model parameters, and efficient handling of sparse data. Its ability to perform parallel computation on

hardware makes it exceptionally fast compared to other standard boosting methods. Furthermore, XGBoost allows for the use of custom optimization objectives and evaluation criteria, adding a layer of customization that can be tailored to specific predictive tasks.

The use of the XGBoost Regressor in predicting educational outcomes illustrates its robust applicability across various settings, underscored by its powerful and efficient processing capabilities. However, the fine-tuning of its parameters—such as depth of trees, learning rate, and number of trees—is crucial to optimizing its effectiveness specific to the dataset at hand. Further studies may explore integrating XGBoost within a broader ensemble or hybrid model framework to capitalize on its strengths while mitigating any potential overfitting or bias seen in singular model applications.

Table 2. The XGBoost Regressor evaluation metrics

Metric	Value
MSE	**41.69**
RMSE	**6.46**
R^2	**0.829**

The XGBoost Classifier (Target variable: letters)

The XGBoost Classifier, standing for eXtreme Gradient Boosting, is a cornerstone in the realm of advanced machine learning techniques used for classification tasks. Its robustness stems from its ability to optimize both computational efficiency and predictive accuracy across large datasets. This model is particularly adept at handling various types of data, making it indispensable for complex analytical challenges in many fields, including education.

The focus of this analysis was the 'Final Grade' assigned to students, an ordinal variable. This means the grades are structured in a meaningful sequence without a fixed interval between them, reflecting their qualitative nature more than quantitative exactness. Such variables are

crucial for tasks where the order impacts outcomes but where precise distances between categories are not defined.

XGBoost builds on the principles of traditional boosting techniques but with significant improvements in terms of execution speed and model performance. It constructs a series of decision trees sequentially, where each tree learns to correct its predecessor's errors, effectively reducing bias and variance. This ensemble method not only enhances accuracy but also incorporates regularization techniques to avoid overfitting—a common pitfall in less sophisticated models.

The application of the XGBoost Classifier in predicting student grades demonstrates its robust capability to adapt to the intricacies of educational data. With its methodical approach to reducing error and its efficient processing of complex inputs, XGBoost stands out as a model of choice for educators and researchers alike. Future investigations might explore further tuning of its hyperparameters or combining it with other machine learning techniques to even better tailor its predictions to specific educational needs.

Table 3. The XGBoost Classifier evaluation metrics (Target variable: letters)

Metric	Value
Accuracy	**76.54%**
Precision	**75.02%**
F1 Score	**75.46%**

The XGBoost Classifier (Target variable: Pass or Fail)

The analysis focused on 'Final Grade' as the target variable, categorized as an ordinal variable within the educational assessment context. Unlike nominal variables, ordinal variables contain inherent ordering but do not presume equal intervals between categories, making them ideal for educational grading systems where such distinctions matter.

XGBoost stands out due to its unique approach to building decision trees in sequence, each designed to correct the errors of its predecessor, thereby enhancing overall prediction accuracy. This sequential building process is supplemented by gradient descent methods to minimize loss functions, effectively reducing bias and variance across predictions. The model's structure includes multiple hyperparameters such as learning rate, number of trees, and max depth, which can be finely tuned to enhance model performance.

Employing the XGBoost Classifier to predict student grades has proven to be highly effective, showcasing the algorithm's robustness and adaptability to complex datasets. Its ability to perform well under diverse conditions suggests it could be further explored in other areas of educational data analysis. Future research might focus on optimizing XGBoost's hyperparameters or integrating it with other machine learning frameworks to further enhance its predictive capabilities.

Table 4. The XGBoost Classifier evaluation metrics (Target variable: Pass or Fail)

Metric	Value
Accuracy	**95.05%**
Precision	**95.09%**
F1 Score	**95.07%**

Model Selection

In the analysis of machine learning models to identify at-risk students based on their academic performance, it becomes imperative to choose a model that not only predicts with high accuracy but also aligns with the objective of the research, which is to distinguish students who pass from those who fail. Based on the comparative results of several algorithms, the XGBoost classifier stands out due to its superior performance metrics.

The results are clear when examining the mean squared error (MSE), root mean squared error (RMSE), and coefficient of determination (R^2). The XGBoost model demonstrates remarkable consistency in its MSE and RMSE values, which are significantly lower than those of the competing Decision Tree Regressor model. The consistency of the error metrics implies that the XGBoost model has a reliable performance with fewer deviations in its predictions, a crucial feature for a robust classification model.

More crucially, the R^2 values for the XGBoost classifier, which indicate the proportion of variance for a dependent variable that's explained by an independent variable or variables in a regression model, are compellingly higher than the other models. Specifically, the R^2 value for the XGBoost model, when the target variable is binary (Pass or Fail), is an impressive 0.9507. This high R^2 value suggests that the XGBoost model can explain a substantial proportion of the variance in the outcome variable, which in this context, refers to the students' ability to pass or fail.

Given the context of the study—to identify at-risk students—the dichotomy of 'Pass' or 'Fail' is of particular importance. It reflects a direct and actionable classification that educational institutions can employ for interventions. The granular grades of A, B, C, D, and F, while informative, may not be as immediately applicable to the goal of early intervention. Thus, a model that can predict with high accuracy whether a student is likely to pass or fail is preferable for implementing preventative measures and support systems.

In conclusion, the XGBoost classifier, with the target variable set to a binary outcome of Pass or Fail, is the most suitable model for this study. It has proven to be the most accurate and reliable in predicting student outcomes, which is crucial for early detection and support of at-risk

students. The implementation of this model will enable educators and administrators to allocate resources effectively, offer timely support to those in need, and ultimately improve the educational attainment of their students. This selection is not only based on statistical reasoning but is also driven by the pragmatic goal of the research, underscoring the applied nature of this master's thesis in addressing real-world educational challenges.

Feature Importance

The evaluation of feature importance is a critical aspect of understanding the predictive power and the underlying behavior of machine learning models. In this study, the XGBoost Classifier's contribution to understanding each feature's effect on predicting student results is analyzed using SHAP (SHapley Additive exPlanations) values. SHAP values help in interpreting the model by quantifying each feature's impact on the prediction outcome, offering both directionality and magnitude.

After fitting the XGBoost model to the training data, SHAP values were computed to capture the impact of each feature on the model's predictions. This approach not only provides insights into which features are most influential but also how different feature values shift the model output from the base value, which is the average model output over the training set.

The SHAP summary plot provides a visual representation of the feature impacts across all test instances. Here are the key observations from the SHAP summary plot:

- High Impact Features: Some features like “EXMin1”, “HWMin1”, and “EXMax1” show significant positive influence on the model's output, indicating that higher values of these features positively correlate with higher probabilities of predicting the positive class. This

suggests that extreme values in exams and homework are critical indicators of student performance.

- Negative Impact Features: On the contrary, features like “QZMin1” and “PQZTrend1” tend to lower the prediction score when they increase. This indicates a negative correlation with the student's likelihood of success, suggesting that lower quiz scores or negative trends in performance metrics might be strong predictors of a student needing intervention.
- Feature Value Distribution: The color coding on the plot, with blue indicating lower feature values and red indicating higher values, further illustrates the nuanced role of these features. For example, high scores in “SAT” and consistent performance in `HWWeight` tend to push the model towards predicting higher student performance.

The analysis of feature importance through SHAP values not only confirms the relevance of traditional academic performance metrics like exams and homework scores but also highlights less obvious factors such as trends over time and quiz scores as significant predictors. This understanding allows educators to focus on holistic student performance monitoring and tailored intervention strategies.

Furthermore, the variability in SHAP values across different features underscores the complexity of educational performance, suggesting that interventions should be multifaceted and personalized based on a range of performance indicators.

The feature importance analysis using SHAP values with the XGBoost Classifier provides a robust framework for identifying key predictors of student success. This approach is

instrumental in developing targeted educational programs and supports, ultimately enhancing educational outcomes by allowing for early identification and support of at-risk students.

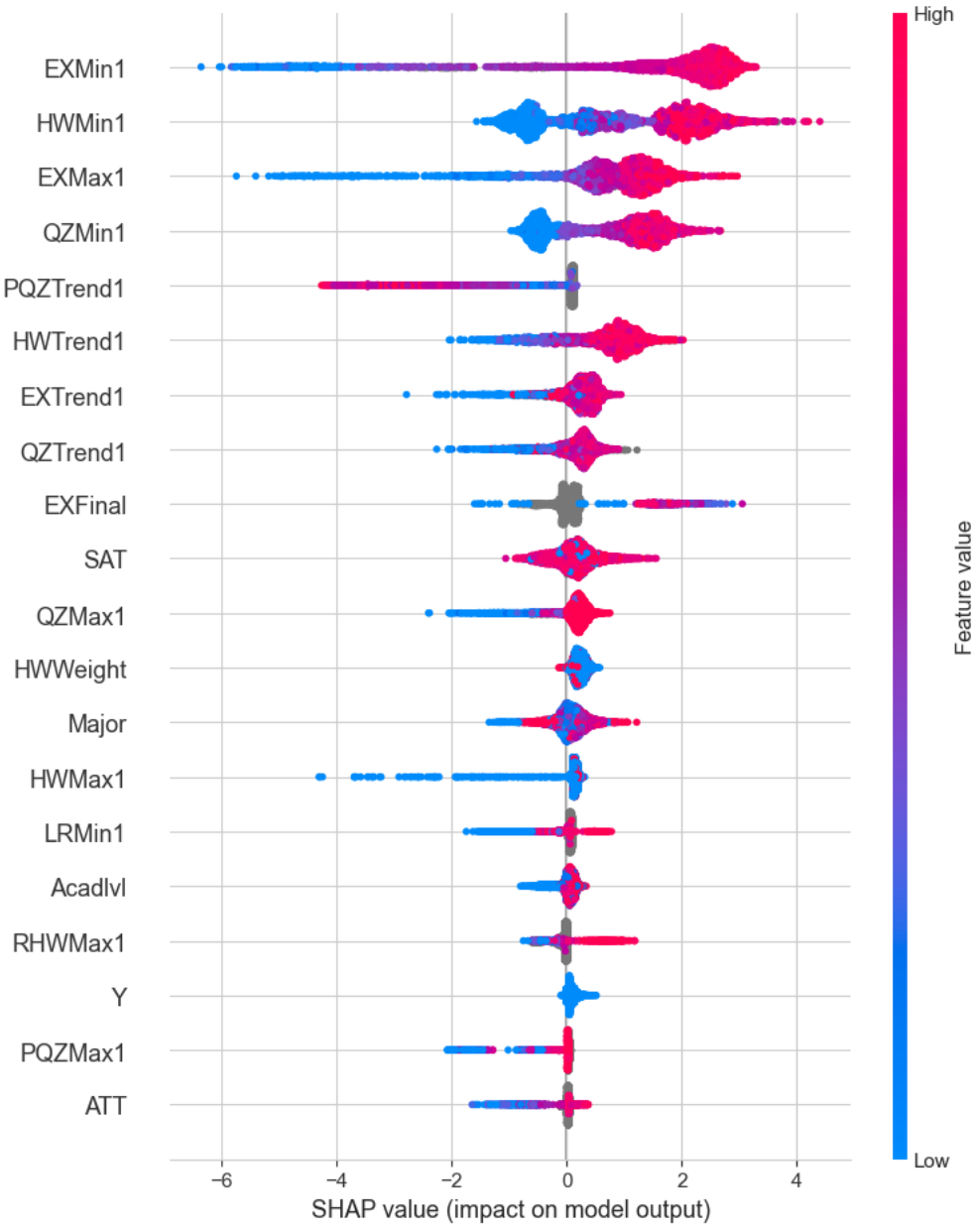

Figure 6. SHAP values showing impact on model output

The SHAP (SHapley Additive exPlanations) waterfall plot provides an insightful way to understand the contribution of each feature to a specific prediction, thereby elucidating how various factors influence student performance.

Using the SHAP waterfall plot, we can decompose a single prediction to observe how each feature contributes towards increasing or decreasing the predicted output relative to the base value, which is the model's average output over the dataset. This approach helps in pinpointing the critical drivers behind each prediction.

In the given example, the predicted final score of the student (f(x)) is 9.815, with the base prediction (E[f(X)]) being 2.965. The plot breaks down the prediction to show how each feature's value influences this score:

- EXMin1: The minimum score in exams significantly boosts the prediction by +2.45, indicating its strong positive impact on performance predictions.
- EXFinal: The final exam score also plays a crucial role, increasing the prediction by +1.96.
- HWTrend1: A positive trend in homework scores over time contributes +1.71 to the prediction, showcasing its importance in predicting improving student performance.
- QZMin1: The minimum quiz score decreases the prediction by -0.4, suggesting that lower quiz scores might indicate a struggle with the subject matter.
- ATT: Interestingly, the attendance (ATT) metric shows a negative contribution of -0.54, which may require further investigation to understand potential reasons behind this negative association.

- Other features such as EXMax1 and SAT scores show moderate positive contributions, while QZTrend1 indicates a slight negative trend, impacting the model's prediction subtly.

This detailed breakdown assists educators and administrators in understanding not just the "what" but the "why" behind predictions of student performance. For instance, the significant positive impact of exam scores suggests that interventions to improve exam preparation could be beneficial, while the negative impact of some quiz scores indicates areas where students might need more support or revised instructional strategies. Furthermore, unusual findings, such as the negative contribution of attendance, warrant a deeper investigation. It may be linked to data anomalies, model biases, or real trends that need addressing through policy changes or targeted interventions. The SHAP waterfall plot serves as an invaluable tool for dissecting model predictions into understandable components, enabling actionable insights. This analysis not only supports transparent decision-making but also helps in fine-tuning educational strategies to cater to individual student needs based on a nuanced understanding of various influencing factors.

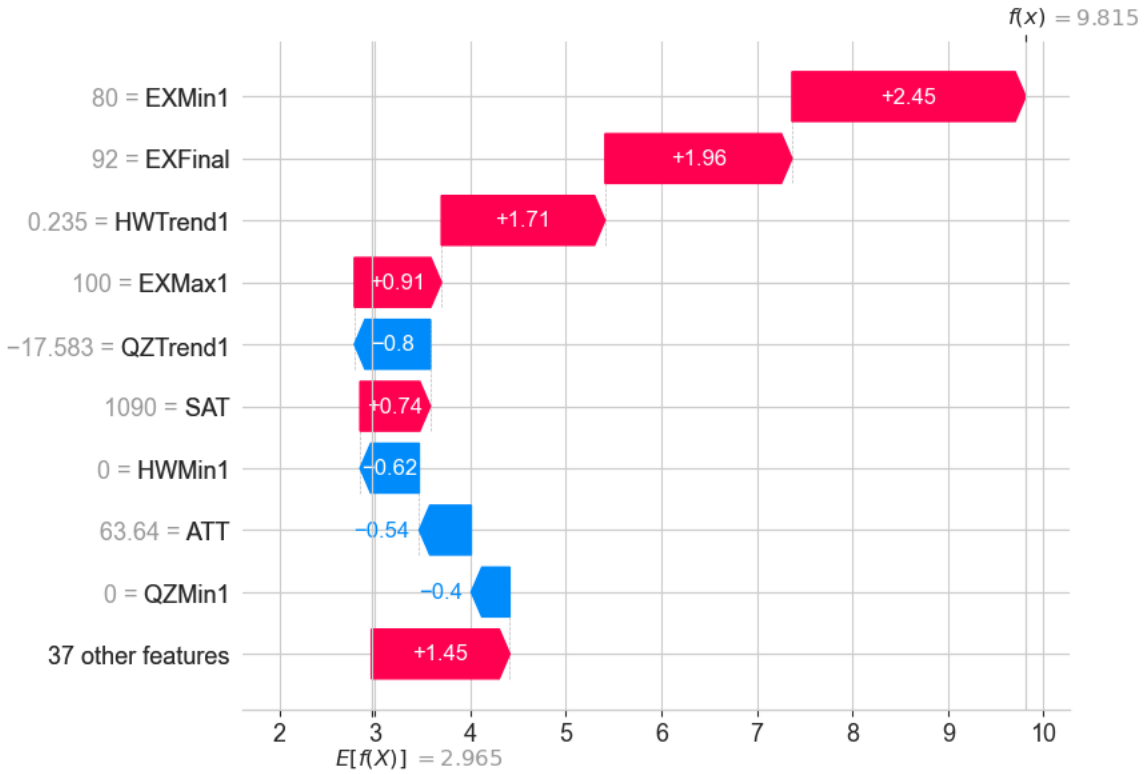

Figure 7. The SHAP waterfall plot

Quantitative assessments such as the F-score from the XGBoost feature importance plot provide a direct measure of each feature's utility in constructing the predictive model. The F-score is derived from how many times a feature is used to split the data across all trees within the model, indicating its relative importance for accuracy.

Using the XGBoost's built-in plot_importance function, we generated a bar chart that ranks features based on their F-scores. This metric effectively captures the frequency of a feature’s usage in making splits in the model's trees, reflecting its significance in improving model accuracy.

The plot reveals several key insights into which features are most influential in predicting student performance:

- HWTrend1: The highest F-score, indicating its pivotal role. This feature tracks the trend in homework scores, suggesting its predictive power in identifying student performance patterns.
- SAT and EXTrend1: Both features also show high importance, underscoring the role standardized tests and trends in exam scores play in educational assessments.
- QZTrend1, EXMax1, and EXMin1: These features, while not as dominant as the top ones, still play significant roles, reflecting the importance of quizzes and exams' variability in assessing student outcomes.
- Features such as ATT (attendance) and demographic factors like Gender and Residency appear less influential. This might indicate that while these factors are relevant, they do not directly impact academic performance as much as academic behaviors and achievements.

The quantitative ranking of features provides a clear hierarchy of what factors should be prioritized when designing interventions or further refining predictive models. For instance, the high importance of homework and exam trends suggests that consistent performance in these areas should be a key focus for educational support programs.

Moreover, the lower importance assigned to attendance could prompt a reevaluation of traditional assumptions about its role in academic success, possibly shifting focus towards more direct measures of student engagement and comprehension.

The combination of quantitative feature importance and qualitative insights from SHAP analysis offers a comprehensive view of the factors driving student performance predictions. This dual approach not only enhances the robustness of the model's interpretations but also

ensures that educational stakeholders can make informed decisions based on a holistic understanding of the data.

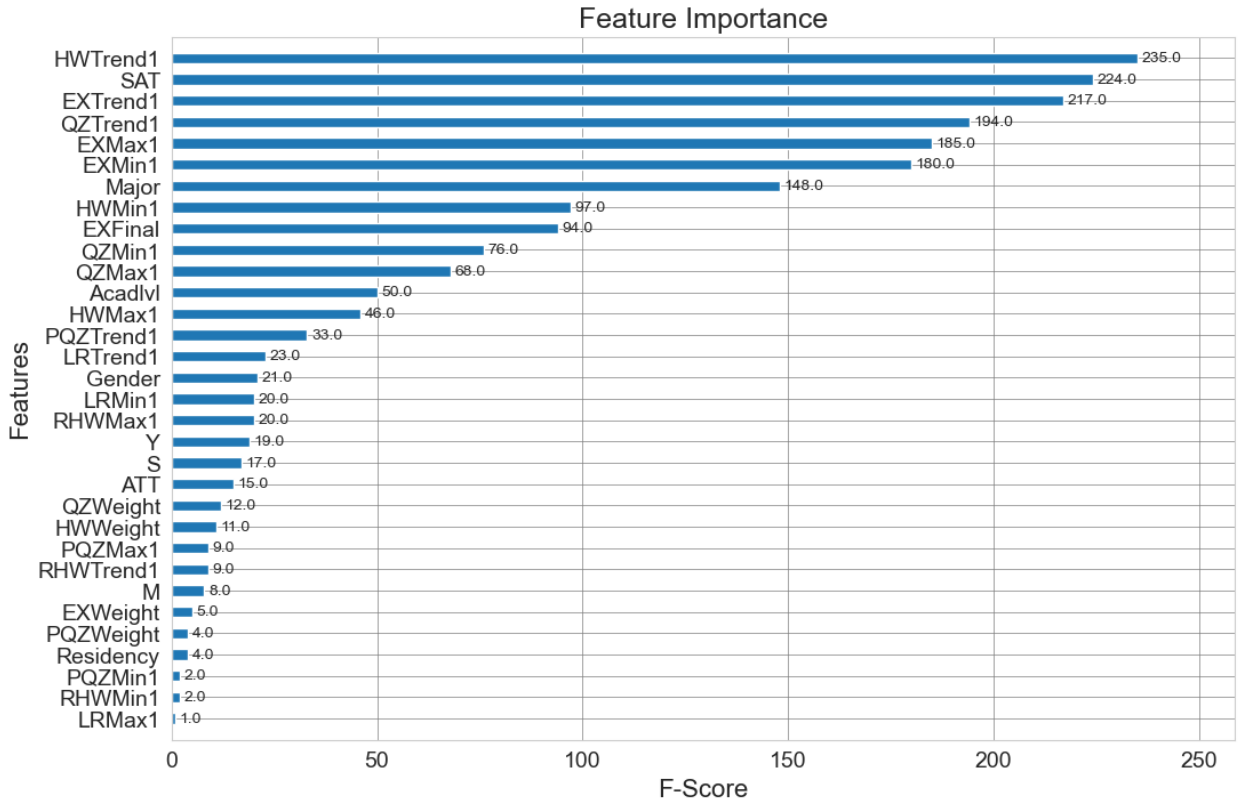

Figure 8. Quantitative Assessment of Feature Contributions and F1 Scores

CHAPTER FIVE: CONCLUSIONS AND SUGGESTIONS

In this chapter, we will analyze the most significant features affecting students' learning outcomes in an introductory statistics course, with the assumption that the final score accurately reflects students' learning outcomes. This examination aims to identify and understand the various factors, including student engagement, teaching methods, assessment techniques, and feedback mechanisms, that significantly influence the effectiveness of the course. ### Conclusion.

Homework Trends

The investigation revealed a direct correlation between positive homework trends and higher median final scores, underscoring the value of consistent improvement and engagement with course materials. Conversely, a decline in homework performance was associated with lower median final scores, highlighting potential challenges in content comprehension or external factors affecting student performance. It should be mentioned that the Homework trends was the most important feature in the model explaining students' final grade in the course. The analysis of students with stable homework scores suggested a consistent, albeit potentially unchallenged, understanding of the material.

The variability in final scores, as indicated by the interquartile range, and the presence of outliers in each category further refined our understanding of the relationship between homework trends and final scores. These elements underscored the diverse ways in which students interact with and benefit from homework assignments, pointing to the necessity of personalized educational strategies.

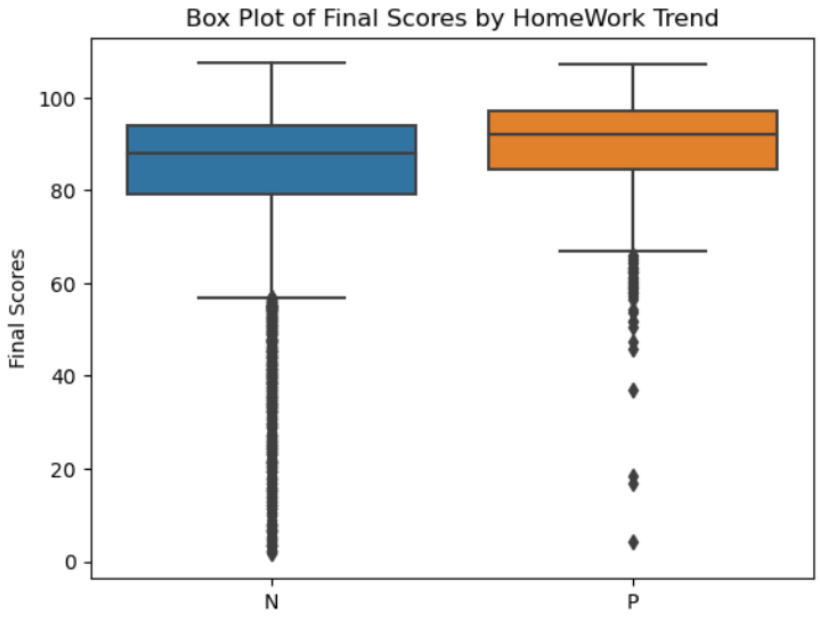

Figure 9. The effect of positive and negative homework trends on students' final scores

Based on the findings, several recommendations emerge for educators and academic institutions:

- Emphasize Regular Practice and Feedback: Reinforce the importance of regular homework completion and provide timely, constructive feedback to support continuous improvement and understanding.
- Early Intervention for Declining Performance: Implement monitoring systems to identify and support students exhibiting a consistent decline in homework performance, addressing potential learning gaps or external challenges early on.
- Encourage Engagement and Exploration: For students exhibiting stable homework performance, introduce more challenging assignments or enrichment activities to stimulate further engagement and academic growth.

- Tailor Support and Intervention Strategies: Utilize data on homework performance trends to develop personalized support mechanisms, catering to the specific needs of students, especially those identified as outliers in performance trends.

The insights gleaned from this analysis advocate for a proactive and differentiated approach to homework assignments, emphasizing the role of educators in guiding and supporting students through their academic journeys. By acknowledging the varied impacts of homework trends on student performance, educators can more effectively foster environments that not only challenge students but also provide the support necessary for them to thrive academically.

In conclusion, the findings underscores the profound impact of homework performance trends on academic achievement.

Exam Trends

The analysis of final scores by exam trends offers a profound insight into the significant variance in students' academic achievements based on their exam performance trends. The classification into positive, negative, and neutral trends not only delineates the students' performance trajectory but also underlines the complex dynamics influencing academic success. The evaluation of these trends reveals essential patterns and outcomes, serving as a foundational basis for our recommendations.

The positive exam trend, characterized by a consistent improvement in scores, correlates strongly with higher final grades. This trend underscores the effectiveness of continuous effort and adaptation in learning strategies. However, the identification of outliers within this group prompts an investigation into the broader spectrum of factors affecting student performance,

suggesting that improvement in exams does not uniformly translate to overall success for every student.

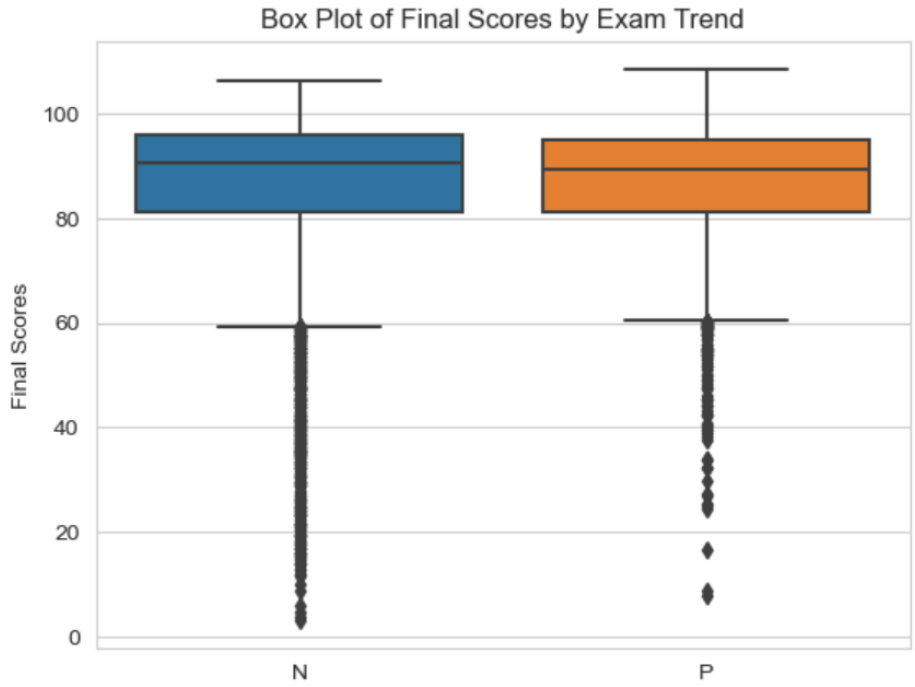

Figure 10. The effect of positive and negative exam trends on students' final scores

Contrastingly, a negative trend in exam scores—indicative of declining performance—highlights potential areas of concern in both curriculum engagement and student well-being. The variability in final outcomes for these students, especially those who manage to offset exam declines with other components, suggests a resilience that merits further support through targeted academic interventions.

The neutral trend group, exhibiting no significant changes in exam performance, provides an interesting lens through which the efficacy of exams in evaluating student understanding can be assessed. This group's performance suggests that stability in exam scores may not necessarily

mean mastery of the material, calling into question the alignment of exams with course objectives.

From an educational standpoint, these insights catalyze a series of recommendations aimed at enhancing teaching methodologies and assessment strategies. Identifying students with declining exam trends for early intervention can significantly mitigate the adverse effects on their final grades. This proactive approach, coupled with the development of personalized learning plans, can address individual learning needs and foster an environment conducive to academic success.

The examination of outliers within these trends advocates for a holistic evaluation framework that integrates multiple forms of assessment. This approach acknowledges the diverse learning styles and external factors influencing student performance, ensuring a fair and comprehensive evaluation system.

In conclusion, analyzing the data emphasizes the critical role of continuous improvement and in student performance. The findings advocate how exam performance trends impact final grades and underscore the importance of adaptive and student-centered teaching and assessment methods. Through targeted support, ongoing feedback, and evaluation, educators can significantly influence students' learning pathways and academic achievements, preparing them for success in their educational paths.

SAT

SAT scores are the second most important features explaining students’ performance in the constructed model. SAT scores, distributed into quartiles from "A" to "D", unfolds a multifaceted view of academic performance within the context of standardized testing. This structured categorization serves as a pivotal foundation for understanding the intricate

relationship between SAT scores and course outcomes, revealing patterns that highlight both the potentials and limitations of standardized tests as predictors of academic success.

The highest quartile (A) showcases a direct correlation between high SAT scores and elevated final course scores, suggesting that students with superior standardized testing capabilities tend to excel academically. This observation supports the conventional wisdom that SAT scores can be a reliable indicator of academic preparedness. However, the presence of outliers within this group prompts a broader consideration of factors influencing student performance, indicating that high SAT scores, while predictive of success, are not the sole determinant of academic outcomes.

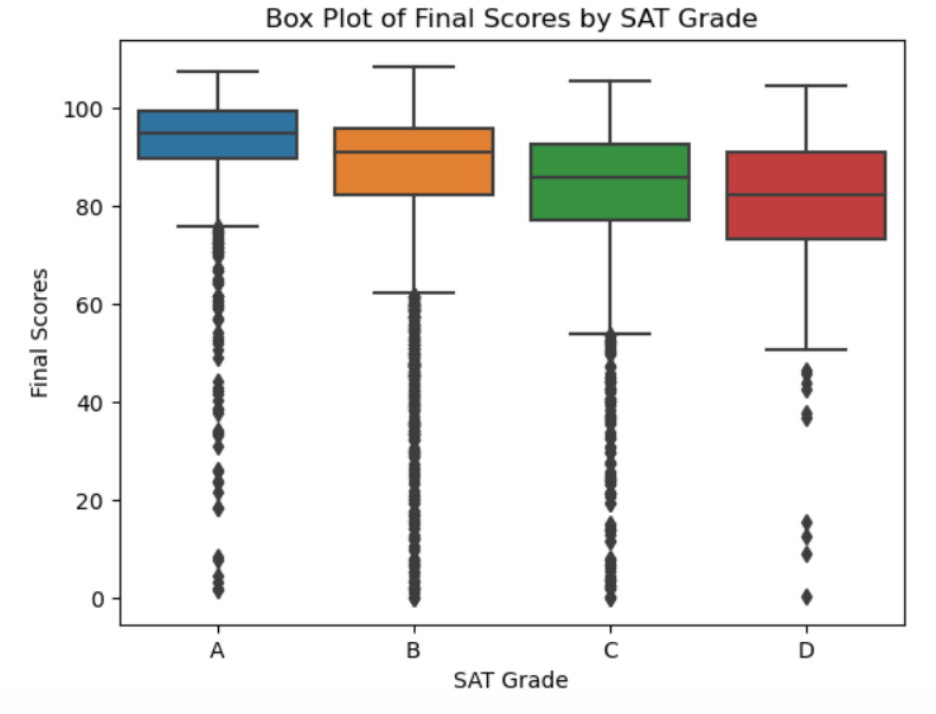

Figure 11. The effect of SAT on students' final scores

Conversely, students in the lower quartiles, particularly "C" and "D", exhibit a wider range of final scores, suggesting a more complex interaction between standardized test scores

and course success. This variability points to the influence of numerous factors beyond SAT scores, including teaching methodologies, student engagement, and individual learning strategies, which can profoundly impact academic performance.

The observations derived from this analysis pave the way for a series of educational implications and recommendations aimed at enhancing teaching and learning experiences. Notably, the need for adaptive learning strategies emerges as a critical theme, underscoring the importance of personalized educational approaches that cater to the diverse needs and capabilities of students across all quartiles. Such strategies could include differentiated instruction, targeted support interventions, and flexible curriculum designs that accommodate varying levels of academic preparedness and learning styles.

Furthermore, the analysis advocates for the adoption of holistic assessment approaches, transcending beyond the confines of standardized testing to encompass a broader spectrum of student abilities and achievements. This entails integrating multiple forms of assessment to capture a comprehensive view of student learning, thereby ensuring a more equitable and inclusive evaluation of academic performance.

Lastly, the significance of supportive interventions cannot be overstated, especially for students who may underperform relative to their SAT quartile. Proactive identification and support for these students can play a pivotal role in mitigating academic challenges and fostering an environment that promotes success for all learners, irrespective of their standardized test scores.

Exam Min Score

The PIs explored the impact of students' poorest performance/scores in the exams (referred to as "Exam Min") on their overall final scores as it was pointed out as an important feature according to SHAP values. The categorization into quartiles based on "Exam Min" scores provided a nuanced insight into how low-performing exams correlate with final academic achievements across different subsets of students. The "Exam Min" scores were categorized into quartiles from A to D, reflecting the range of scores from highest to lowest within the minimal scoring exams. This classification was derived from a function applied to the "EXMin1" scores, where:

- Quartile A represents scores from 75 to 100,
- Quartile B from 50 to less than 75,
- Quartile C from 25 to less than 50,
- Quartile D from 0 to less than 25.

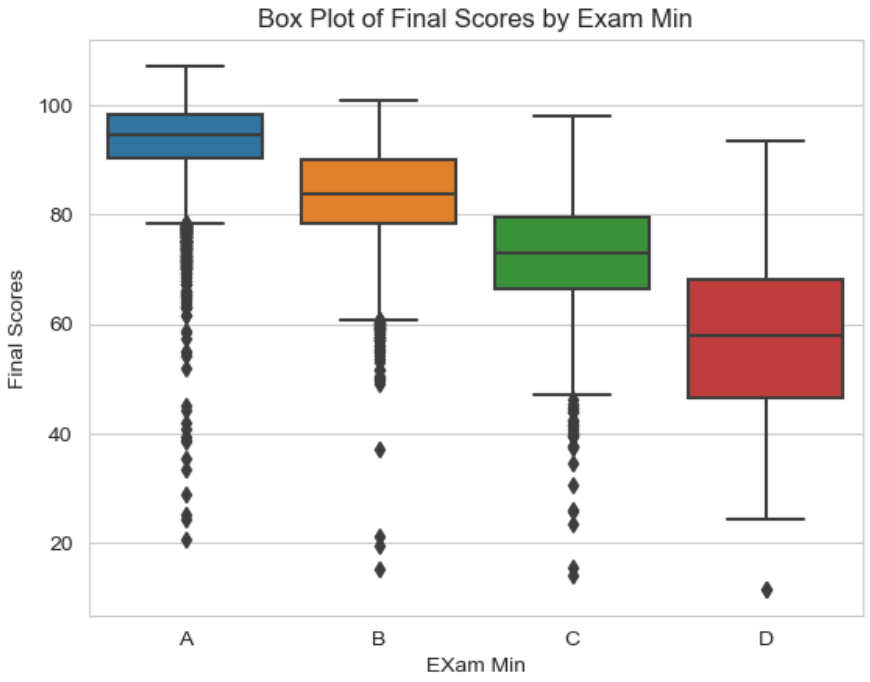

Figure 12. The effect of minimum score in exams on students' final scores

This segmentation allows for a focused analysis on how the lowest scores in a student's profile influence their overall performance. The box plot (Figure ?) illustrates the distribution of final scores across the four Exam Min quartiles:

- Quartile A (Blue Box): Students in this quartile generally achieve the highest final scores, with a narrow interquartile range (IQR) indicating less variability among students. This quartile also exhibits fewer outliers, suggesting a strong correlation between high minimum exam scores and overall high final scores.
- Quartile B (Orange Box): This group shows a wider IQR, reflecting greater variability in final scores. The presence of outliers below the lower whisker indicates that some students, despite higher minimal scores, may struggle to achieve comparably high final scores.
- Quartile C (Green Box): The median final score drops significantly in this quartile, with an even wider IQR. This suggests a less consistent correlation between Exam Min scores and final outcomes, with considerable variation in student performance.
- Quartile D (Red Box): Exhibiting the lowest median score and the widest spread, this quartile confirms the negative impact of very low minimum exam scores on final grades. The numerous outliers both above and below the box highlight individual variations, where some students still manage to surpass expectations despite low scores in one or more exams.

The analysis clearly demonstrates that higher scores in the lowest-performing exams are indicative of better overall academic performance. Conversely, students who score particularly

low in any exam are at risk of lower final scores, though exceptions exist as indicated by the outliers.

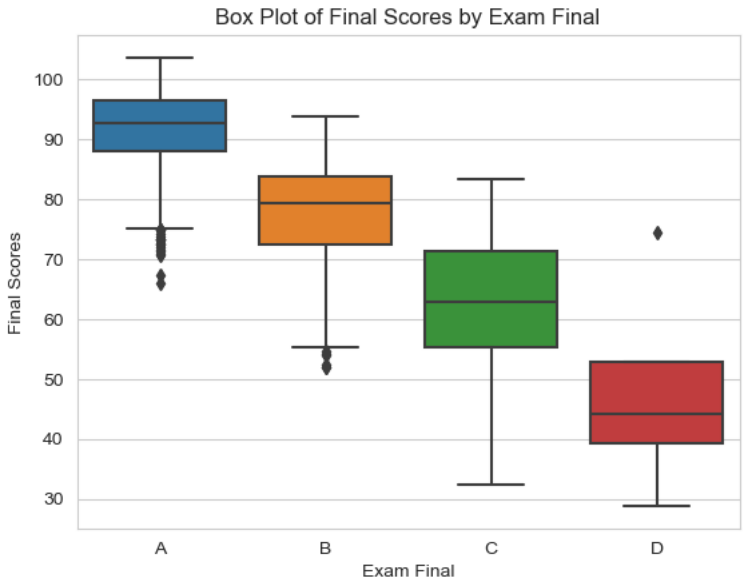

Figure 13. The effect of the final exam on students' final scores

Exam Final Scores

In this section, we investigate the impact of students' performance in their final exams, categorized into quartiles, on their overall final scores. The "Exam Final" scores were segmented into four quartiles, A, B, C, and D, reflecting ascending ranges of final exam scores. The quartiles were determined using the following ranges:

- Quartile A: 75 to 100 (highest scores),
- Quartile B: 50 to less than 75,
- Quartile C: 25 to less than 50,
- Quartile D: 0 to less than 25 (lowest scores).

This quartile system serves to categorize students based on their final exam performance and correlate these categories with their overall course final scores. The box plot provided (Figure ?) offers a detailed breakdown of final scores by exam final quartiles:

- Quartile A (Blue Box): Students in this highest quartile typically achieve the best final scores, as indicated by the higher median and more compact interquartile range (IQR). This quartile shows fewer outliers, reinforcing the correlation between high final exam scores and overall high course performance.
- Quartile B (Orange Box): This quartile displays a slightly lower median final score and a wider IQR compared to Quartile A, suggesting some variability in how well students perform overall, despite reasonably good exam scores.
- Quartile C (Green Box): There is a noticeable drop in the median final score and an even greater spread in scores. The presence of outliers and a broad IQR indicate significant variability, suggesting that while some students manage decent overall performance, others struggle considerably.
- Quartile D (Red Box): Representing the lowest exam performers, this quartile has the lowest median final score and exhibits a wide range of final scores, with several outliers indicating that a few students perform either much better or much worse than the median. This variability underscores the critical impact of poor performance on final exams.

The analysis underscores the critical influence of final exam performance on overall course outcomes. Students scoring higher on final exams tend to secure better overall grades, highlighting the exam's role as a significant determinant of academic success in the course.

Homework Trends and SAT

The section tries to explain the intricate relationships between SAT scores, homework trends, and final course scores, offering a rich tapestry of insights into the dynamics of student performance. Through the analytical lens of SAT quartiles juxtaposed with homework trend signs, this study has unraveled the nuanced ways in which these variables interplay to shape academic outcomes. The categorization of SAT scores into quartiles—A through D—and the classification of homework trends into positive ("P"), negative ("N"), or neutral signifies a methodical approach to dissecting the multifaceted nature of academic success.

The analysis reveals a striking synergy between high SAT scores and positive homework trends, where students exhibiting both characteristics tend to achieve the highest final scores. This correlation underscores the premise that students with inherent academic aptitude, as suggested by their SAT scores, can further enhance their academic trajectories through diligent engagement and continuous improvement in their coursework. Conversely, the study illuminates the significant impact of homework trends on students within lower SAT quartiles, where a positive homework trend can dramatically uplift academic performance, transcending the limitations potentially implied by lower standardized test scores.

Moreover, the observation of students with neutral homework trends across all SAT quartiles offers a pivotal insight into the complex matrix of factors influencing student performance. It suggests that while consistency in homework may maintain a baseline academic performance, it is the trajectory of improvement or decline in homework engagement that bears a more definitive influence on final course outcomes.

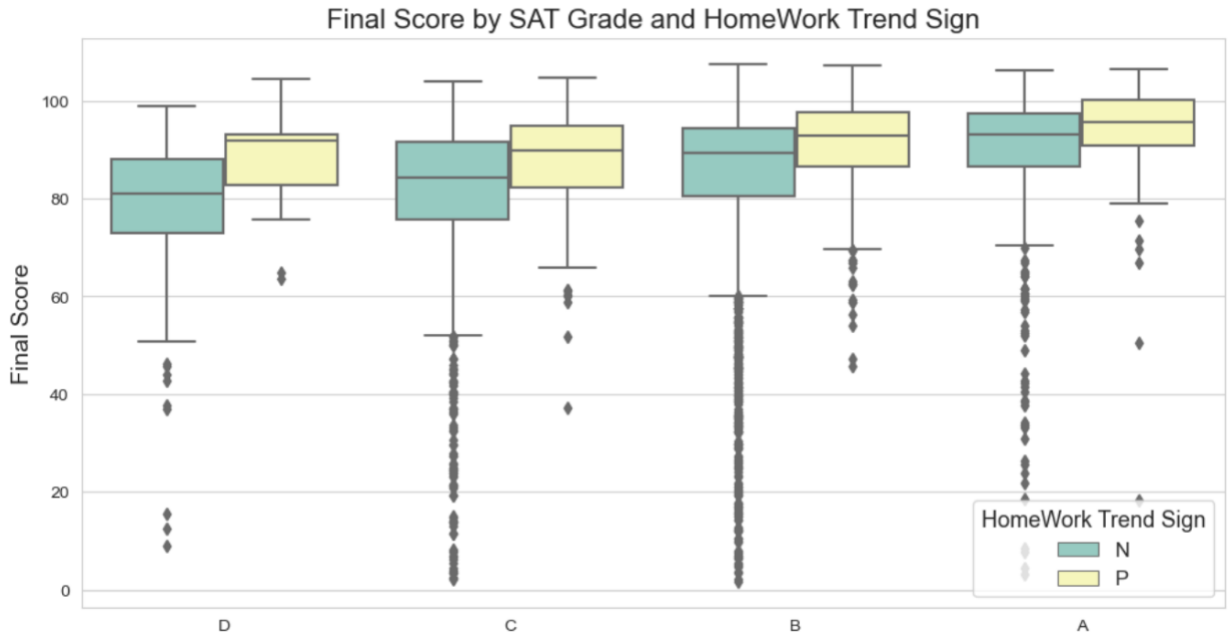

Figure 14. Final score by SAT and Homework trend sign

From these insights emerge several cogent recommendations aimed at fostering an educational ethos that nurtures student success across diverse spectra of baseline academic preparedness and engagement:

- Emphasizing Consistent Engagement and Improvement: Educators are encouraged to cultivate an environment that values and rewards regular homework completion and continuous improvement, recognizing these behaviors as pivotal drivers of academic achievement.
- Personalized Support Systems: The study advocates for the implementation of personalized support mechanisms, particularly for students in lower SAT quartiles or those exhibiting negative homework trends. Initiatives such as tutoring, study groups, and bespoke homework assistance programs are recommended to address individual learning gaps and foster a supportive academic environment.

- Assessment Methods: In light of the findings, there is a pressing need for holistic assessment methodologies that encompass a broader spectrum of student performance indicators. These should extend beyond standardized test scores to include measures of engagement, improvement, and effort in coursework, offering a more rounded evaluation of student achievement.

It can be concluded that through the adoption of personalized, holistic, and adaptive educational strategies, institutions can more effectively support their students in reaching their utmost academic potential, irrespective of their starting points. This approach not only democratizes academic success but also prepares students for lifelong learning and adaptation in an ever-evolving world.

Exam Trends and SAT

Throughout this analytical process, the exploration into the interplay between SAT scores, exam trend signs, and final course scores has unveiled complex layers of student academic performance. This analysis, segmenting SAT scores into quartiles from "A" to "D" and correlating these with exam trend signs (indicating improvement "P", decline "N", or stability), has provided a nuanced perspective on the multifaceted nature of academic achievement. This dual-focused investigation enriches our understanding of the predictive value of standardized tests alongside the crucial influence of exam performance trends on students' final academic outcomes.

The intricate relationship revealed through this analysis underscores a significant revelation: students in the highest SAT quartiles who also show a positive trend in exam scores typically achieve the highest final scores. This synergy between inherent academic aptitude and a

trajectory of improvement in exam performance suggests a robust pathway to academic excellence. Conversely, observations within the lower SAT quartiles, especially among students manifesting a positive exam trend, challenge the conventional reliance on standardized testing as the sole predictor of academic success. This demographic, despite lower SAT scores, often achieves commendably high final scores, spotlighting the transformative potential of consistent improvement in exam performance.

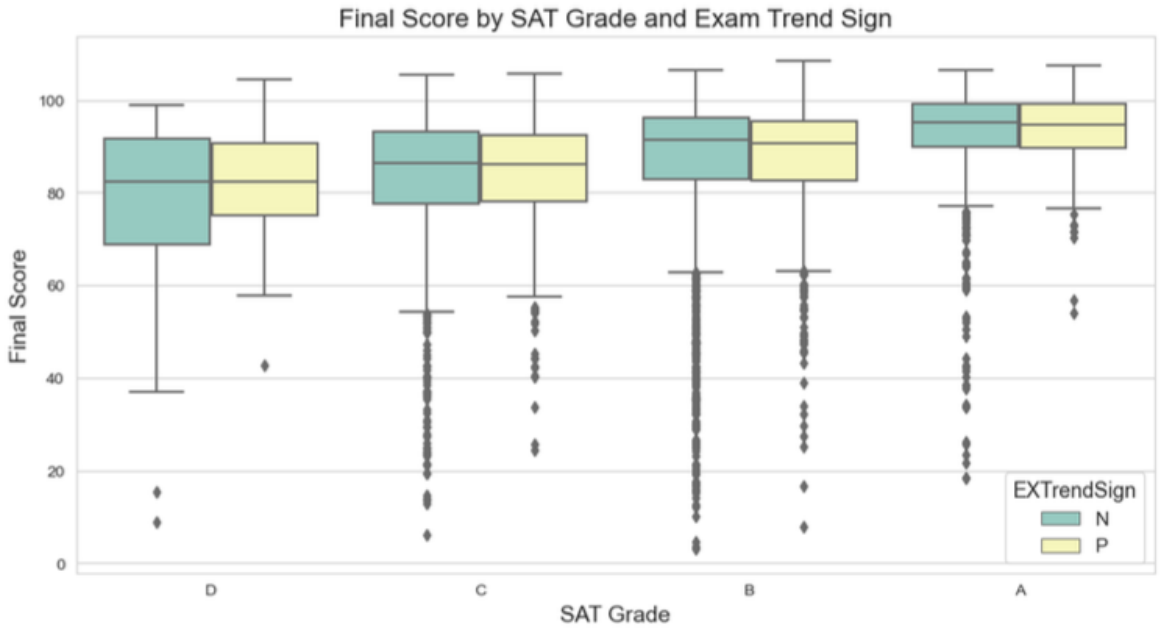

Figure 15. Final score by SAT and exam trend sign

Furthermore, the analysis brings to light the diverse academic outcomes of students exhibiting a neutral exam trend across all SAT quartiles, pointing towards a balanced, yet intricate, relationship between exam stability and final academic success. It suggests that while stability in exam performance does not inherently detract from achievement, the directionality of exam performance—whether improvement or decline—more accurately predicts final course outcomes.

Conclusion

In this chapter, we explored several key factors that influence student performance in an introductory statistics course. Each factor contributes uniquely to the academic outcomes of students and understanding their interplay is crucial for developing effective educational strategies. Here are detailed insights into the major factors discussed:

Homework performance emerged as a pivotal indicator of student success. Our analysis delineated three distinct trends:

- Positive Trends: Students showing consistent improvement in their homework scores generally achieved higher final scores. This correlation suggests that regular engagement and mastery over course content significantly boost overall performance.
- Negative Trends: A decline in homework performance was linked to lower final scores, indicating struggles with course material or external factors affecting student performance.
- Stable Trends: Students with stable homework scores often demonstrated a consistent understanding, though this might also suggest a lack of challenge in the assignments.

These trends underline the necessity of regular practice and timely feedback, highlighting that educators should focus on reinforcing homework as an integral part of learning, not just as a formality.

Exam performance trends provide deep insights into students' learning trajectories over the course:

- Positive Trends: Improvement in exam scores from one assessment to the next consistently correlated with higher overall final grades. This pattern underscores the effectiveness of adaptive learning strategies and the importance of resilience and effort.
- Negative Trends: Declining exam scores highlighted potential areas of concern such as engagement issues or gaps in understanding, necessitating early intervention and targeted support to reverse these declines.
- Neutral Trends: Students whose exam scores did not significantly vary posed questions about the exams' alignment with course objectives, suggesting that stability in scores might not always equate to mastery of the material.

These insights advocate for a holistic evaluation of exams as both reflective of and influential to students' academic performance.

The analysis of SAT scores segmented into quartiles revealed:

- High Quartiles (A and B): Higher SAT scores were generally associated with better course performance, supporting the view that standardized tests can predict academic preparedness and success.
- Low Quartiles (C and D): A broader range of final scores in these quartiles indicated that SAT scores are not the sole predictors of success, emphasizing the impact of other factors like teaching methodologies and student engagement.

This distribution suggests that while high SAT scores indicate potential, they must be complemented by robust educational support to ensure that all students can achieve their academic goals.

Exam Min Scores refer to the lowest scores students achieved in any of their exams throughout the course. This measure is particularly telling as it highlights the areas where students struggle the most. Analyzing these scores provided insight into:

- Quartile A (Highest Scores): Students with higher minimal exam scores generally showed higher overall performance, suggesting that consistent competency across exams correlates strongly with better final outcomes.
- Quartile D (Lowest Scores): Conversely, those with very low minimum scores often had lower final scores. This pattern highlights the negative impact that failing to grasp key concepts in any part of the course can have on overall performance.
- Impact of Support: The analysis suggests that targeted academic support, especially for students struggling with specific exams, can significantly improve their understanding and performance in subsequent assessments.

These insights stress the importance of continuous performance monitoring and the provision of timely academic interventions to help students overcome their weakest areas.

Exam Final Scores reflect students' performance on their final exam, often a significant component of their overall course grade. This score is crucial because it encapsulates a student's cumulative understanding of the course material. The analysis segmented these scores into quartiles, revealing:

- Quartile A (Highest Scores): Students in the highest quartile for final exam scores typically also had high overall course scores, underscoring the final exam's weight and the efficacy of mastering course content.

- Quartile D (Lowest Scores): Those in the lowest quartile often struggled significantly, impacting their overall grades adversely. This demonstrates the critical nature of the final exam in determining final course outcomes.
- Need for Comprehensive Preparation: The importance of preparing students throughout the semester for this culminating assessment is evident, as it significantly influences their final grades.

These findings emphasize the final exam's role as a critical determinant of academic success and advocate for strategies that ensure all students are adequately prepared for this decisive assessment.

Both Exam Min and Exam Final scores are essential indicators of student performance, each providing unique insights into different aspects of their academic journey. Exam Min scores help identify knowledge gaps early on, while Exam Final scores assess cumulative knowledge and readiness for course completion. Together, they underscore the need for a balanced approach to education that addresses both immediate and comprehensive learning outcomes.

Based on these factors, our recommendations focus on enhancing educational practices to support diverse learning needs:

- **Proactive Support:** Recognizing and addressing negative performance trends early can help mitigate their impact. This involves more than just academic support; it includes counseling and mentoring to address external factors affecting student performance. Students in lower quartiles might benefit from additional academic support and resources to help improve their understanding and performance in challenging areas.

- **Monitor Performance Closely:** Identifying students who perform poorly on certain exams can help in intervening early, potentially improving their outcomes in subsequent assessments.
- **Continuous Assessment and Feedback:** Implementing continuous assessment mechanisms can help identify and support students at risk of falling into lower performance quartiles before the final exam.
- **Engagement and Enrichment:** For students displaying stable or positive trends, introducing challenging materials and enrichment activities can further enhance their engagement and academic growth.
- **Personalized Education:** Tailoring teaching methods and assessments to accommodate diverse student needs ensures that all students can benefit from the educational system.
- **Encourage Consistent Performance Across Exams:** Educational programs should aim for a balanced approach where students are equally prepared across different exams, reducing the risk of low scores in any single assessment impacting the overall grade significantly.

This analysis forms an integral part of our broader study on factors influencing student performance, highlighting the importance of minimal exam scores as a significant predictor of academic success. These strategies are designed to foster an environment that not only challenges students but also supports them through their academic journey, ensuring no student is left behind due to predefined academic metrics or unaddressed learning gaps.

APPENDIX A: INSTITUTIONAL REVIEW BOARD FORMS

UCF IRB Letter

UNIVERSITY OF CENTRAL FLORIDA

Institutional Review Board
FWA00000351
IRB00001138, IRB00012110
Office of Research
12201 Research Parkway
Orlando, FL 32826-3246

NOT HUMAN RESEARCH DETERMINATION

November 29, 2023

Dear Shahab Abbaspour Tazehkand:

On 11/29/2023, the IRB reviewed the following protocol:

Type of Review:	Initial Study
Title of Study:	Predicting Undergraduate Students' Scores in a Statistical Introductory Course Using Machine Learning Algorithms
Investigator:	Shahab Abbaspour Tazehkand
IRB ID:	STUDY00005914
Funding:	None
Documents Reviewed:	• HRP-251 - FORM - Faculty _SignedAdvisor Scientific-Scholarly Review.pdf, Category: Faculty Research Approval; • HRP 250, Category: IRB Protocol; • OtherAtttachments, Category: Other;

The IRB determined that the proposed activity is not research involving human subjects as defined by DHHS and FDA regulations.

IRB review and approval by this organization is not required. This determination applies only to the activities described in the IRB submission and does not apply should changes outside of administrative ones (study personnel, timelines, etc.) be made. If non-administrative changes are made (design, information collected, instrumentation, funding, etc.) and there are questions about whether these activities are research involving human in which the organization is engaged, please submit a new request to the IRB for a determination by **clicking Create Modification / CR** within the study.

If you have any questions, please contact the UCF IRB at 407-823-2901 or irb@ucf.edu. Please include your project title and IRB number in all correspondence with this office.

Sincerely,

Tamiko Fukuda
UCF IRB

REFERENCES

Abbaspour, S. (2022, July). Supporting Secondary Teachers' Proof and Justification of Calculus Concepts Through the Intentional Use of Dynamic Technology. University of Central Florida Electronic Theses and Dissertations, 2020-. 1166. https://stars.library.ucf.edu/etd2020/1166

Abbaspour, S., & Safi, F. (2023, October) The Role of Beliefs, Visualization and Technology in Teaching and Learning Proof: The Case of Skylar, Proceedings of the 45th Annual Meeting of The North American Chapter of the International Group for the Psychology of Mathematics Education. Reno, NV.

Abbaspour, S., & Safi, F. (2021, October). Infusing Proof and Justification, Mathematical Modeling and Technology: The Case of Mathematical Series. Proceedings of the 43rd annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education. Philadelphia, PA.

Alyahyan, E., & Düştegör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, *17*(1), 3.

Bergstra, J., Bardenet, R., Bengio, Y., & Kégl, B. (2011). Algorithms for Hyper-Parameter Optimization. Proceedings of the 24th International Conference on Neural Information Processing Systems, 2546-2554.

Bhutto, E. S., Siddiqui, I. F., Arain, Q. A., & Anwar, M. (2020, February). Predicting students' academic performance through supervised machine learning. In *2020 International Conference on Information Science and Communication Technology (ICISCT)* (pp. 1-6). IEEE.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794. DOI: 10.1145/2939672.2939785.

Evans, J. H., Whigham, P. A., & Wang, L. (1995). The Influence of Role Models on Aspiring Scientists. Educational Studies in Mathematics, 29(3), 259-271.

GE, O., Mamah, C. H., Ukekwe, E. C., & Nwagwu, H. C. (2020). A machine learning based framework for predicting student's academic performance. *Physical Science & Biophysics Journal*, *4*(2).

Lykourentzou, I., Giannoukos, I., Nikolopoulos, V., Mpardis, G., & Loumos, V. (2009). Dropout prediction in e-learning courses through the combination of machine learning techniques. *Computers & Education*, *53*(3), 950-965.

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. Frontiers in Neurorobotics, 7(21). DOI: 10.3389/fnbot.2013.00021.

Shahiri, A. M., & Husain, W. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science*, *72*, 414-422.

Wang, D., Lian, D., Xing, Y., Dong, S., Sun, X., & Yu, J. (2022). Analysis and prediction of influencing factors of college student achievement based on machine learning. *Frontiers in Psychology*, *13*, 881859.

Wang, L., Whigham, P. A., & Evans, J. H. (2009). Student Evaluations of Instruction: Data Mining Approach. Journal of Educational Data Mining, 1(1), 52-71.

Xu, J., Zhao, L., & Liu, X. (2018). Classification of Gene Expressions in Cancer Patients Using XGBoost Algorithm. Journal of Medical Systems, 42(11), 204. DOI: 10.1007/s10916-018-1044-1.

Zabriskie, C., Yang, J., DeVore, S., & Stewart, J. (2019). Using machine learning to predict physics course outcomes. *Physical Review Physics Education Research*, *15*(2), 020120.

Zhao, L., & Hryniewicki, M. K. (2018). XGBoost Applied to Fraud Detection: Implementation and Interpretation. Journal of Financial Crime, 25(4), 984-1003. DOI: 10.1108/JFC-04-2018-0043.