Summer 2023

Campus Safety Data Gathering, Classification, and Ranking Based on Clery-Act Reports

Walaa F. Abo Elenin

Follow this and additional works at: https://digitalcommons.georgiasouthern.edu/etd

Part of the Databases and Information Systems Commons, Data Science Commons, and the Information Security Commons

Recommended Citation


This thesis (open access) is brought to you for free and open access by the Jack N. Averitt College of Graduate Studies at Digital Commons@Georgia Southern. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons@Georgia Southern. For more information, please contact digitalcommons@georgiasouthern.edu.
ABSTRACT

Most existing campus safety rankings are based on criminal incident history with minimal or no consideration of campus security conditions and standard safety measures. Campus safety information published by universities/colleges is usually conceptual/qualitative and not quantitative and are based-on criminal records of these campuses. Thus, no explicit and trusted ranking method for these campuses considers the level of compliance with the standard safety measures. A quantitative safety measure is important to compare different campuses easily and to learn about specific campus safety conditions.

In this thesis, we utilize Clery-Act reports of campuses to automatically analyze their safety conditions and generate a safety rank based on these reports. We first provide a survey of campus safety and security measures. We utilize our survey results to provide an automated data-gathering method for capturing standard campus safety data from Clery-act reports. We then utilize the collected information to classify existing campuses based on their safety conditions. Our research model is also capable to predict the safety rank of campuses based on their Clery-Act report by comparing it to existing Clery-Act reports of other campuses and reported rank on public resources.

Our research on this thesis uses a number of languages, tools, and technologies such as Python, shell scripts, text conversion, data mining, spreadsheets, and others. We provide a detailed description of our research work on this topic, explain our research methodology, and finally describe our findings and results. This research contributes to the automated campus safety data generation, classification, and ranking.

INDEX WORDS: Campus safety, Data extraction, Dataset generation, Safety classification, Safety and security ranking, Data mining, Regression analysis.
CAMPUS SAFETY DATA GATHERING, CLASSIFICATION, AND RANKING 
BASED ON CLERY-ACT REPORTS

by

WALAA ABO ELENIN
B.S., Alderson Broaddus University, 2019

A Thesis Submitted to the Graduate Faculty of Georgia Southern University
in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE IN INFORMATION TECHNOLOGY
COLLEGE OF ENGINEERING AND COMPUTING
CAMPUS SAFETY DATA GATHERING, CLASSIFICATION, AND RANKING
BASED ON CLERY-ACT REPORTS

by
WALAA ABO ELENIN

Major Professor: Lei Chen
Committee: Yiming Ji
           Jongyeop Kim

Electronic Version Approved:
July 2023
DEDICATION

I would also like to dedicate this work to my daughter Jana, for bringing joy and hopes to me and my family. I also dedicate it to my son Kareem for always being so hopeful and motivated. I dedicate this work also to my son Amr for always taking the lead and initiative to explore new success horizons. I also dedicate this work to my husband Atef for his constructive advice and wonderful support. I sincerely would like to thank my research supervisor Dr. Lei Chen for all the support to complete my master’s thesis. I also would like to thank Dr. Yiming Ji and Dr. Jongyeop Kim for their wonderful support and their encouragement. I am also grateful for their effort to review and enhance my thesis dissertation.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................................................. 4
LIST OF FIGURES .................................................................................................................................................. 5
CHAPTER 1 - INTRODUCTION - CAMPUS SAFETY MEASURES AND FACTORS ............................................. 6
  1.1 Introduction to Campus Safety ...................................................................................................................... 6
  1.2 Campus Safety Components, Groups, and Measures ................................................................................... 8
    1.2.1 Campus Safety Components and Measures .......................................................................................... 9
    1.2.2 Campus Safety Ranking ....................................................................................................................... 21
  1.3 Statistical Analysis of Campuses Safety Component and Measures ......................................................... 22
  1.4 Summary of Campus Safety Components and Measures .......................................................................... 26
CHAPTER 2 - CAMPUS SAFETY MODELS AND PROTECTION METHODOLOGY ........................................ 27
  2.1 Campus Safety Models and Methodology .................................................................................................... 27
  2.2 Literature Review of Existing Campus Safety Mechanisms ...................................................................... 29
    2.2.1 Classification of Existing Campus Safety Models .............................................................................. 32
    2.2.2 Campus Safety Methods ....................................................................................................................... 32
    2.2.3 Campus Safety Environments, Goals, and Concerns ......................................................................... 39
  2.3 Campus Phishing Attack Detection and Protection Methods ........................................................................ 43
    2.3.1 Results and Findings ............................................................................................................................. 55
  2.4 Summary of Campus Safety Models and Protection Methods .................................................................... 56
CHAPTER 3 - DATASET GENERATION FOR CAMPUS SAFETY RANKING .................................................... 57
  3.1 Introduction .................................................................................................................................................. 57
  3.2 Background and Related Work .................................................................................................................... 58
  3.3 Data Gathering Mechanism ....................................................................................................................... 60
    3.3.1 Clery-Act report definition and Structure .......................................................................................... 61
    3.3.2 Campus Safety Ranking Data .............................................................................................................. 64
  3.4 Data Extraction Model Overview .............................................................................................................. 65
  3.5 Input Data Preparation ................................................................................................................................ 66
  3.6 Specifying Extracted Data Structure .......................................................................................................... 67
  3.7 Data Extraction Model Implementation .................................................................................................... 69
  3.8 Sample Campus Safety Dataset Results .................................................................................................... 70
  3.9 Summary of Dataset Generation for Campus Safety Ranking .................................................................... 74
CHAPTER 4 - AUTOMATED CAMPUS SAFETY CLASSIFICATION AND RANKING ...................................... 75
  4.1 Model Overview ....................................................................................................................................... 76
  4.2 Classification Model ................................................................................................................................... 77
    4.1 Campus Safety Classification Results ......................................................................................................... 79
    4.2 Campus Safety Ranking Model .................................................................................................................. 82
    4.3 Campus Safety Ranking Results .............................................................................................................. 84
    4.4 Summary of Campus Safety Classification and Ranking ........................................................................ 90
CHAPTER 5 - CONCLUSIONS ............................................................................................................................ 91
REFERENCES ..................................................................................................................................................... 94
APPENDIX A Python code for dataset generator ............................................................................................... 98
LIST OF TABLES

Table 1- Main safety goals, concerns, methods, and target environments in the current research. ........... 30
Table 2- Anti-Phishing Solutions .............................................................................................................. 46
Table 3- Sample Clery-Act report contents - Western Michigan University Clery-Act report 2021 ........... 63
Table 4- Sample of campus safety ranking for a list of universities and colleges across the United States 64
Table 5- Statistical data measures used in our classification process. ...................................................... 77
Table 6- Classification model calculation sample data from different safety classes. ........................... 80
Table 7- Prediction precision by regression model for different campus safety classes. ....................... 84
LIST OF FIGURES

Figure 1- Main classes of campus safety and security measurement.................................................. 8
Figure 2- Main subclasses of campus safety and security measurement ........................................... 10
Figure 3- Safety Management of campus safety and security measurement .................................... 11
Figure 4- Safety and the Security Systems of campus safety and Security Measurement ................. 15
Figure 5- Crime and Offence Prevention of campus safety and security measurement ..................... 17
Figure 6- On campus Safety Facilities of campus safety and security measurement ...................... 19
Figure 7- Campus Safety Ranking according to campus safety and security measurement .............. 21
Figure 8- Number of campuses per Safety Index ............................................................................. 22
Figure 9- Number of campuses per safety rank .............................................................................. 23
Figure 10- Number of Campuses per Campus Safety classes and Subclasses .................................. 24
Figure 11- Top classes of campus safety ranks .............................................................................. 25
Figure 12- Low classes of campus safety ranks .............................................................................. 25
Figure 13- Relationships of main campus safety goals, concerns, methods, and target environments .. 31
Figure 14- Number of techniques utilized in campus safety .............................................................. 32
Figure 15- “Same Phishing URLs accessing results between Safari and Safari with OCR solution” [56]. 52
Figure 16- Data Extraction Model Overview .................................................................................. 61
Figure 19- Data extraction from Clery-Act reports and dataset generation model .............................. 65
Figure 20- Pdftotxt command sample .............................................................................................. 66
Figure 21- Sample shell script for converting multiple pdfs into text files ....................................... 66
Figure 22- Sample output of campus Clery-Act report Pdfs to Text files using Windows shell script... 67
Figure 23- Safety term groups and aliases ....................................................................................... 68
Figure 24- Campus safety data extraction and dataset generation model overview ............................ 69
Figure 25- Main functions of Python implementation of Campus Safety dataset from Clery-Act reports. 70
Figure 26- Sample dataset based on Safety term occurrence counter in Campus Clery-Act Reports .... 71
Figure 27- Sample dataset based on Safety term existence counter in Campus Clery-Act Reports ........ 72
Figure 28- Sample dataset prepared automatically and structured for Weka data mining tools .......... 73
Figure 29- A sample dataset automatically created for Weka when opened on the Weka editor ........ 73
Figure 30- Campus safety classification model overview ................................................................. 77
Figure 31- Campus safety classification model details ....................................................................... 79
Figure 32- Three campus safety classes based on campus safety data and current ranking index ....... 79
Figure 33- Number of Campuses per safety class by Accepted Density Margin ............................. 81
Figure 34- Campus safety prediction model overview ..................................................................... 82
Figure 35- Campus safety prediction model implementation .............................................................. 82
Figure 36- Sample WEKA-ready (.arff) data files for different classes and different ADM values ....... 83
Figure 37- Prediction precision by regression model for all campus data ........................................ 85
Figure 38- Prediction precision by regression model for class 0 with ADM = 3 ................................. 85
Figure 39- Prediction precision by regression model for class (-1) with ADM = 3 ............................... 86
Figure 40- Prediction precision by regression model for class 1 with ADM = 3 ................................. 86
Figure 41- Correlation Coefficient by prediction model by class with ADM = 3 ............................... 87
Figure 42- Total Number of campuses and number of Instances used for regression for each ADM ...... 87
Figure 43- Gaussian Process Regression by ADM for Class 0 .......................................................... 88
Figure 44- Linear regression by ADM for Class 0 ............................................................................ 88
Figure 45- Multilayer Perception regression by ADM for Class 0 ..................................................... 89
Figure 46- Simple Linear Regression by ADM for Class 0 .............................................................. 89
Figure 47- Random Forest Regression by ADM for Class 0 ............................................................ 90
CHAPTER 1
INTRODUCTION - CAMPUS SAFETY MEASURES AND FACTORS

Protecting students, residents, teachers, and faculty in schools, universities, colleges, and is the ultimate goal for these institutes. A safe campus environment can assist in preventing violence of all kinds in the university area, thus supporting a safe, productive educational environment for everyone, which can lead to high productivity. One effective mechanism to prevent violence on the campus ground is to comply with the campus safety and security rules and regulations designed according to standardized safety and security measures concerning campus support. This research paper focus on the main component of the Clery-Act that is utilized for campus safety and security measurement, strategies preparation for higher institutions’ practice toward safe campus, and other facts that contribute to increased safety and security.

1.1 Introduction to Campus Safety

Choosing the right college or university after high school education is challenging for students and their families. This decision can be accomplished according to a variety of factors, such as geographic considerations, financial support, and other safety measures, such as environmental safety, campus safety, and security. The higher education act 1965 (HEA) is a federal law that governs the administration of higher education programs. The higher education act intends to strengthen the educational resources of universities and colleges in the United States. The Act required colleges and universities that are participating in students’ financial assistance programs to release campus violation facts and security information.

The higher education act was modified and renamed the law Jeanne Clery Disclosure of campus security policy and crime statistics act. Clery-Act reports are the US's most important campus safety reports. Most US universities and colleges prepare an annual report for their safety standards and crime incidents during the reporting interval. This allows tracking campus safety evolution and fulfillment with
the standard practices. Clery-Act reports include qualitative information illustrating safety and security policies and procedures. Thus, despite being rich in information, these reports cannot be an easy source of information for students, parents, and families to learn about campus safety. A quantitative safety measure is important to compare different campuses easily and to learn about specific campus safety conditions. The Clery Act is based on criminal accidents on campus, referred to as the Clery Act. Clery Act reports include information about the violence against women act (VAWA), stalking, sexual assault, hate crime, domestic violence, and several other crimes. Academic institutions are obliged to reveal the implemented strategic plan and programs that institutions utilize to prevent these types of violence [1].

The goal of this thesis is to study and evaluate various methodologies that contribute to campus safety. We first describe a survey of main campus safety components, groups, and measures. We then provide a literature review of campus safety techniques and their goals, safety concerns, and application environments. We emphasize phishing attacks targeted at students and workers in educational environments. We also provide a data gathering and dataset generation technique for extracting campus safety data from Clery-Act reports based on our taxonomy in this Chapter. We utilize our campus safety data generator to classify campuses based on their safety conditions, public ranking, and neighborhood influence on their public rank. Our campus safety classification technique sets the stage for our campus ranking method that can control the level of neighborhood influence on safety ranking while considering the campus safety conditions of individual campuses.

The techniques proposed in this thesis are introduced in the following chapters as follows. Chapter 2 provides surveys of campus safety models and protection methodology in the current research literature. Chapter 3 describes a dataset generation method for campus safety ranking from Clery Act reports annually published by academic institutes. Chapter 4 details our automated campus safety classification and ranking techniques, and Chapter 5 provides our conclusions.
In the rest of this chapter, we describe the main campus safety measures and factors. We describe these factors and categorize them into campus safety components and subcomponents (or groups). We also describe the existing campus safety ranking methods and provide some statistical analysis and discussion on our taxonomy.

1.2 Campus Safety Components, Groups, and Measures

Here we aim to categorize into a group that can be utilized as a standard for universities and colleges to increase safety and security. The classification of the safety and security factors is collected for the USA Universities annual security act, which is utilized as the main component for camp take safety and security measures. The applied techniques are used for statistical analysis to rank the campus safety and security based on the universities’ practices of the applied safety and security mechanisms. The campus safety rank classification takes different formats such as on-campus safety rank, off-campus safety rank, and overall safety rank.

In this section, we aim to study the main classes and subclasses of the campus safety and security components which utilized a major factor in determining the safety and security rank of a university’s campus. Then we illustrate the main classes and subclass components employed for evaluating campus safety and security, including safety measurement, safety, and security systems, crime and offense prevention, on-campus safety facilities, and the campus safety ranking [2].

![Figure 1 - Main classes of campus safety and security measurement.](image)
1.2.1 Campus Safety Components and Measures

In this literature review, we classify the safety and security measurement into five measure aspects: which includes 1) Safety management, 2) safety and security systems, 3) crime offense prevention, 4) on-campus safety facilities, and 5) campus safety ranking. In the next section, we illustrate these components and divide them into subcomponents or groups where each group includes a number of campus safety measures or factors.

Figure 1- Main classes of campus safety and security measurement. represents the main classes of the campus safety and security classification. This includes the major concerns that were used in the Clery-Act specification. In the following, we discuss these key components, plot them in a taxonomy of campus safety components on the top of the tree, and measures at the leaves. The main components of our taxonomy shown above are defined below.

The safety management component contains the following elements annual security report, emergency management, security policy, and safety education programs. These systematic tactics encompass policy and education, which contribute to managing risk related to safety. As an example, the policy of emergency notification system, missing student policy, and personal safety training to increase awareness.

The safety and security system class includes the flowing element cybersecurity, campus safety technology, and building security which comprises different means and devices to protect people and campus facilities from any danger encountered or risks. For example, the utilized devices include camera surveillance, safety mobile applications, GPS tracking, access control, and security training for individuals to be secure from cyber threats.

Crime and offense prevention class include the following elements crime prevention and sexual violence, which address various policy and procedure in terms of crime prevention. Examples of prevention methods encompass policy prohibiting harmful devices or equipment on campus, policies protesting violence against women, and hate crimes.
On-Campus safety facilities include the following elements: environmental health and safety, fire prevention, and transportation safety, which contribute to increasing the safety of campus facilities and people by applying different mechanisms such as chemical lab safety, food safety, fire detection, safe walk programs for individual protection, etc.

Finally, the campus Safety ranking class includes the on-campus safety index and on-campus safety rank which is used to measure campus safety and security based on the university’s practices of the realistic campus safety factors.

Figure 2 - Main subclasses of campus safety and security measurement.

Figure 2 includes the subsections of the safety measurement; I will describe the main subsection under each main class of the safety and safety categorization. In the following subsubsections, we explain the details of our taxonomy and define each safety subcomponent and measures in this hierarchy.
1.2.1.1 Component 1: Safety Management

As shown in Figure 3, the safety management group encompasses the following four measure components of the annual security report, emergency management, security policy, and safety education programs. Each component contains elements that describe the procedure and the techniques implemented to maintain a secure and safe campus.

![Figure 3- Safety Management of campus safety and security measurement.](image)

1.2.1.1.1 Campus Annual Security Report

To receive federal funding, US academic institute campuses must annually provide a security report based on Clery-Act specifications. The report contains information about security policy, crime statistics, procedures, fire statistics reports, and security programming and resources. It also contains
information about the applied universities and colleges’ policies and mechanisms to manage risk related to safety and security. To receive federal funding, universities, and colleges must disclose clear information about certain crimes in their annual reports. Institutions that do not comply with the regulation of the Clery Act can be subject to a financial fine. The annual security report contains about three years of campus crime statistics information, and it is designed to provide essential information about the safety and security of university campuses [3].

The Clery Act identifies certain institutional employees as campus security authorities (CSAs). Under the Clery Act, these people must notify the university’s police of the specific crime. These individuals are mandated to report crimes that happen on campus, witness, and/or report them. CSAs are required to report the type of crime, the location of the crime scene, when it happened, and when it was reported to them.

1.2.1.1.2 Emergency Management

Campus emergency management is a framework that includes managerial functions, rules, regulations, policies, and procedures that are utilized to reduce safety vulnerabilities, hazards, and campus-related risks. Emergency management includes emergency drills, response guides, calls, crisis intervention teams, and notifications. Emergency drills and practices are essential for organizations. They prepare employees and organizations for actual emergency situations. The practice can be used for fire and other emergencies.

Campus emergency phones are distributed throughout campuses for emergency contact and are connected to university public safety. These phones automatically display the caller’s location for individuals’ safety. Clery-Act demands that higher education institutions disclose their information about the applied policies and procedures for emergency notification to campus employees and students for significant emergency threats that might affect their health and safety. The emergency response guide is designed to provide necessary instruction on on-campus emergencies and how to respond to emergencies
on the university’s campus. A crisis intervention team program intends to provide urgent intervention to individuals in a mental health crisis and provide awareness to support them [4].

1.2.1.1.3 Security policy

Security policy refers to the related organization’s policy, practice, and procedures that are utilized to assist organizations to protect against threats. Higher educational institutions are required to publish several safety and security policy statements that are related to crime prevention Policy, Alcohol Policy, drug abuse violations, smoking policies, safety policies, and missing student policies which are applied to missing students who reside in on-campus housing. The following section presents an illustration of the campus security policy.

The crime prevention policy program is a set of educational rules and regulations that aim to protect students and minimize crime that can take place on campus. Crime prevention policy focuses on the safety of individuals and addresses various safety-related concerns such as safety on the street, safety on public transportation, theft prevention, safety in residences, fire safety, etc. The drug and alcohol policy required Institutions under the Clery-Act to participate in the federal student financial aid program must have a drug and alcohol abuse prevention program accessible to employees and students. Thousands of U.S. colleges and universities practice different policies that apply to their culture to prevent any harm associated with drugs and alcohol to protect individuals. Smoke policy objectives are supporting individuals from exposure to secondhand smoke and keeping the university environment smoke-free for students, staff, and faculty with a comfortable and healthy environment [5].

Student code of conduct is a set of rules, principles, and exceptions that are given to students and their parents that outline students’ responsibilities and rights to ensure that students’ behavior meets the university’s expectations. Missing student policy describes that the institution must issue a policy statement that addresses missing student notification and the applied procedures that the institution will follow if any students are determined to be missing within 24 hours. The incident of the missing student must be reported to the campus police, security team, or local law enforcement agency that oversees the
area. The missing student policy statement should include a list of titles, such as the organizations or persons. Students are encouraged to register one or a few people to be contacted in case the person is missing [1]. The Department of public safety supports the universities mission by implementing different techniques and safety procedures to provide a safe campus environment for everyone. Public safety officers enforce rules and regulations on campus and maintain a strong relationship with the police department, fire department, and faculty, staff, and students to achieve the universities mission [6].

1.2.1.1.4 Safety education programs

Universities’ safety educational programs aim to enhance the safety of staff, faculty, and students. Safety educational programs are part of the institution’s educational mission, intended to reduce the chance of risks and hazards individuals can face in the future. The Education safety programs can include personal safety, environmental safety, and safety against active campus violence. As an example, Ohio State University provides personal safety programs to empower and teach students, faculty, and staff different mechanisms for self-protection.

Self-defense training begins with educational awareness, prevention, and risk reduction and avoidance approaches. Health & Safety online training is also provided, and it is required based on the employee’s work environment and considers the employees' specific hazards might experience as a part of their jobs. Safety against active campus violence is part of the Ohio State University safety educational program, which aims to prepare individuals to respond to an active, aggressive campus situation and help people to avoid panic and disorientation [7].
1.2.1.2 Component 2: Safety and Security Systems

Figure 4- Safety and the Security Systems of campus safety and Security Measurement.

Figure 4 Safety and security systems utilize technology and security training to maintain a safe campus and ensure the safety of individuals. Safety and security systems include elements such as cybersecurity, campus safety technology, and building security, which protects the universities and colleges’ systems from a broad range of hazards. In the following section, we illustrate the safety and security systems elements.

1.2.1.2.1 Cybersecurity Training for Higher Educational Institutions

Security awareness training for employees, faculty, and administration is essential to protecting the organization’s system and network from cyberattacks. It helps to identify and respond appropriately to possible threats. Security awareness training assists in providing employees with the knowledge needed to identify, report, and prevent cybersecurity-related risks. Cybersecurity threats can be phishing attacks, data breaches, social engineering, ransomware, malware, and other malicious activities. Organizational
cyber security training can assist in mitigating the risks associated with cyber-attacks and ensure the protection of an organization’s sensitive data and system. In addition to the security training, universities apply more mechanisms such as secure computer and device registration to increase the security of individuals’ devices around the campus. These safety precautions include keeping records of individuals’ equipment that include a description of the equipment along with the make, model, and serial number for tracking devices [8].

1.2.1.2.2 Campus Safety Technology

Universities and colleges utilize several Hi-tech techniques to maintain a safe and secure campus environment for everybody. The campus security system is an effective and mandatory mechanism to achieve a strong foundation of safety to protect students, faculty, staff, campus assets, and campus resources. Applied effective security systems on campus ground can guide campus officers to prevent instances of violence. It is challenging to find a reliable, comprehensive security system that works efficiently for the entire campus and covers all the safety and security concerns of a university. Most universities use the advancement of technology and utilize integrated security systems related to the technology and comply with their demand. The integrated security system can include elements such as Safety Mobile App, GPS Tracking, Secure Communication, Camera Surveillance, etc. [9].

The applied mechanism is utilized as a user based on safety and security concerns. As an example, Safety mobile apps are utilized by universities colleges students for self-protection to make students safer and more secure. Some of students' most used safety mobile applications include Circle of 6, LiveSafe, Companion, On Watch, Sky App, Campus Safe, Drunk Mode, etc. LiveSafe application can also benefit students to connect to their family and friends to increase safety precautions. Furthermore, GPS tracking systems are an active approach to tracking people’s location and guiding them in emergencies quickly located [10].
Security cameras are considered a foundation of the security system for universities and colleges and are considered an effective part of the security system. The benefits of security cameras include but are not limited to improving emergency preparedness and response, reducing theft and potential threats, reducing bullying, and increasing the environmental safety of individuals. A security camera can monitor and record the area of concern and create a live video that can be shared with the emergency responders, which helps to proactively detect suspicious behavior and guide campus police to take the required action at the right time [11].

1.2.1.2.3 Building Security

Building access control is an applied technique by some universities and colleges to create a secure environment. Building security is an important practice for universities and colleges utilized to supervise and maintain the accessibility of the university’s facilities. Building security includes access control, student housing, and lecture hall security. Access control systems enable universities and colleges to monitor people entering and exiting buildings, residence halls, and lecture halls. Access cards, pin codes, key cards, or any other access control methods, and keeping out unauthorized persons on the property. Several universities apply various security measures for students’ protection. For example, campus key cards are used on some campuses for students to access campus. If the students’ access card is lost or stolen, students can identify security instantly to disable their card and prevent unauthorized people from accessing campus buildings or any campus-restricted area [12].

1.2.1.3 Component 3: Crime and Offense Prevention

<table>
<thead>
<tr>
<th>Crime and Offence Prevention</th>
<th>Sexual Violence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crime prevention</td>
<td>Sexual Consultation</td>
</tr>
<tr>
<td>Active Shooter</td>
<td>Sexual Assault</td>
</tr>
<tr>
<td>Firearms</td>
<td>Violence Against Women</td>
</tr>
<tr>
<td>Weapon</td>
<td>Staking</td>
</tr>
<tr>
<td>Hate Crime</td>
<td></td>
</tr>
<tr>
<td>Suicide</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 5- Crime and Offence Prevention of campus safety and security measurement.*
Figure 5 Crime and offense prevention includes crime prevention and sexual violence, which encompasses various prevention methods utilized by the campus to prevent the campus from violence, sexual violence offense, and sexual consultation. In the following section, we illustrate the crime and offense prevention elements.

1.2.1.3.1 Crime prevention

That can include special training, policy, applied rules, and regulations to protect the campus ground from any violence. The policy includes prohibiting weapons, firearms, cutting instruments, and other harmful tools on campus. The crime prevention techniques include active shooter training for university staff, students, and other university members to spread awareness among individuals. Clery-Act requires colleges and universities to report any crime under the category of hate crime and spread awareness through an educational program among individuals to prevent this type of crime. Hate crime is defined according to the Clery-Act handbook as “a criminal offense that manifests evidence that the victim was intentionally selected because of the perpetrator’s bias against the victim”. Hate crime bias can be race, sexual orientation, religion, gender identity, national origin, ethnicity, and disability. The suicide prevention team aims to assist students seeking help to prevent the destruction oneself crime [13].

1.2.1.3.2 Sexual Violence

Sexual violence is a wide term that encompasses sexual assaults, domestic violence, dating violence, and stalking. Sexual violence has serious negative and physical effects on colleges students. Violence Against Women Act (VAWA) is a law that provides comprehensive support to any violence agist women. On March 7, 2013, the (VAWA) Act amendments to the Clery Act, and these amendments required some changes which require institutions to disclose programs, policies, and statistics related to dating violence, sexual assault, domestic violence, and stalking. These changes require higher education institutions to disclose the implemented programs used to prevent these types of violence (domestic
violence, dating violence, stalking, and sexual assault) on campus. Most Universities provide consultation to assist individuals who have experienced sexual violence [1] [14].

1.2.1.4 Component 4: On-Campus Safety Facilities

Figure 6- On campus Safety Facilities of campus safety and security measurement.

Figure 6 includes environmental health and safety, fire prevention, and transportation safety. Environmental health and safety consist of various safety measures such as food safety, chemical lab safety, environmental safety, and weather safety. Fire prevention includes a smoke detector and fire prevention practices. Transportation safety includes safety-related practices, which include bike registration and safe walking programs.

1.2.1.4.1 Environmental Health and Safety (EHS)
Environmental health and safety are practices that aim to maintain safety and quality in the work environment. Universities and colleges put effort into applying procedures to identify hazards and keep the work environment safe. Environmental health and safety encompass different safety aspects that can involve safety related to the environment, food safety, chemical safety, and weather-related safety (EHS). Food Safety includes the inspection of the provided food on campus and ensuring that food provided on campus is safe for consumption.

The Environmental Health of universities is responsible for following up on reports of food-borne illness and managing the campus food services in large events such as campus conferences [15]. Campus chemical lab safety is a set of standards and procedures of safety roles for individuals that use the chemical lab to promote safety in laboratories. Campus weather safety. Campus weather safety practices help maintain a safe environment for students, staff, and faculty and reduce weather-related risks. Universities maintain emergency alerts to notify individuals of severe weather changes and update them with the campus safety procedure [16].

1.2.1.4.2 Fire Prevention

Fire prevention aims to increase awareness, recognize the hazards, and address the necessary and correct action in fire emergencies. Some Universities and colleges provide fire safety training for students and organizations. During the training, they highlight fire hazards and concerns and address the effectiveness and the necessity of fire and smoke detection in different campus areas. The U.S. fire administration collects and provides some documents for college students that can be used for outreach activities to increase awareness and keep students safe from the fire on and off campus. The provided safety tips address issues, fire risks, and prevention actions that can prevent students from harming themselves. Universities and colleges are required to disclose safety information to publish and report an annual fire safety report to students and employees, which contains the applied university’s policies statements as well as fire statistics related to on-campus students' housing, including the number of fire incidents reported, injuries, the reason that caused the fire, etc. [17].
1.2.1.4.3 Transportation Safety

Transportation can be defined as the techniques students utilize to manage their rides to and from campus. Transportation takes the form of campus transit, biking, walking, etc. Some universities provide bike registration programs to secure individuals’ bikes against theft. For example, the University of Houston encourages faculty, staff, and students to register their bikes with the university-related service. Upon registration, the university will come up with a serial number that is associated with the bicycle, and this information can guide the university services to return the bicycle to the real owner in case of being stolen. Since the safety of students is a measurable concern for various universities, universities exercise safe ride programs for students, staff, and faculty available 24 hours/7 days a week [18][19].

1.2.2 Campus Safety Ranking

![Campus Safety Ranking](image)

*Figure 7- Campus Safety Ranking according to campus safety and security measurement.*

Figure 7 campus safety ranking includes the following two categories on-campus safety index and on-campus safety rank.

Numerous safety factors and measurements are utilized to evaluate campus safety tracking. The campus safety and Security Clery-Act illustrates and the most significant factors and measurements utilized to examine the campus crime statistics in detail, which are required as important and effective components of the safety measurement. Universities that participate in federal student aid programs are required to release the annual security and safety report according to the Clery-Act measurement, and these reports are used to measure the university safety ranking according to the released results. The campus safety rank can be measured based on the campus safety index, off-campus safety index, overall
safety index, campus safety rank, on-campus safety rank, and overall safety rank. Our analysis focuses on the campus safety index and on-campus safety ranking as the main key measurement for campus safety [20].

1.3 Statistical Analysis of Campuses Safety Component and Measures

In this section, we provide the classes and subclasses of existing campus safety techniques and illustrate the scope of these classes, including the number of campuses per index, number of campuses per rank, and percentage of campuses utilizing the safety factors and subfactors.

![Number of Campuses per Safety Index](image)

Figure 8-Number of Campuses per Safety Index.

Figure 8: first, the Chart presents the number of campuses per rank class, divided into 5 segmentations that include rank 1, rank 2, rank 3, rank 4, and rank 5. Each class presents the number of universities recognized within the class rank out of 141 selected universities. The breakdown of the number of campuses per rank class is divided as follows: Rank one equal 16, rank two equal 27, rank three equal 32, rank four equal 33, and Rank 5 equals most of the campuses at the higher-level safety
indexes are equally distributed over different safety index classes. The least number of campuses exist in the safety index class 0, these are the campuses that consider minimal effort towards their safety.

![NUMBER OF CAMPUS PER RANK](image)

*Figure 9- Number of campuses per safety rank.*

Figure 9 divides the number of campuses per rank into ten classes for the selected number of universities which includes 141 campuses. Campus rank is classified in the range of 0 to 900, and each class indicates the number of campuses per rank. For example, the rank class 0 to 100 includes 19 universities, the rank for class 100 to 200 includes 16 campuses, and the rank class 100 to 200 includes 16 campuses. Small classes ranked 0 to 700 include an almost equal number of campuses, and the higher class ranked 800 to 1000 include a very small number of campuses. This shows the distinctions of the higher safety campuses concerning their efforts to become safe campuses.
The sunburst in Figure 10 displays details about the number of universities practicing the safety and security measurement component and subcomponent. The first two layers of the sunburst are similar to levels two and three of the campus safety measure taxonomy structure in Figure 2. In addition to that, the sunburst classifies the percentage of universities out of the 141 utilized simply that are applying the campus safety factors and subfactors. For example, safety management is divided into three categories: annual security, emergency management, security policy, and education programs. Each of these categories is divided into subcategories, and the sunburst figure signifies the number of universities that practice these subgroups, for example, emergency management including emergency response guide, and
105 out of 141 universities practice the emergency management techniques and 137 practice the emergency notification system for safety management.

Figure 11 - Top classes of campus safety ranks.

Figure 12 - Low classes of campus safety ranks.

In chart Figure 11 and Figure 12, we contrast the low-level safety rank classes versus high-level safety rank classes concerning the number of campuses supporting each safety measure per rank class. The x-axis is the various campus safety measures, and the y-axis is the number of campuses supporting these measures in each safety rank class. For example, the five low-level classes have a maximum of 18 campuses per class practicing emergency drills, while the top-level classes have a maximum of 52 campuses per class.
practicing the same safety measure. The Y-axis scale of the top-level classes (60 campuses) is higher than that of the low-level classes (20 campuses). In these two charts, we also notice that most college campuses in both low-level and top-level classes pay less attention to cybersecurity and safety technology measures: security training, secure computers, device registration, mobile safety app, GPS tracking, and secure communication. The only safety measure in campus safety technology measures that have high attention on these campuses is the camera surveillance systems.

1.4 Summary of Campus Safety Components and Measures

Campus Safety and security are important factors in maintaining a highly productive environment for everyone including students, faculty, and staff. The government implements some rules and regulations to ensure the university’s campus is safe and secure for everyone, including the annual security report. Colleges and universities that receive federal funding are required to disseminate annual security reports to participating in the financial aid program are required to submit the annual security report for staff and students. The annual security reports include statistics of campus crime and details about the techniques and procedures universities apply to increase and enhance campus safety. In this research, we pointed out the main factors that can be utilized to increase campus safety based on the annual security report of universities, and the university’s ranking is determined according to the highest safest environment with little or no crime.
CHAPTER 2

CAMPUS SAFETY MODELS AND PROTECTION METHODOLOGY

Campus safety alerting and awareness systems have continuously improved by utilizing new methods and technologies to assure students’ and workers’ safety. Current research introduces numerous safety techniques to support an individual’s safety during the day and night lives on campuses. These techniques provide safety support in different and specific environments or based on specific safety concerns. Thus, several safety threats are usually unsupported on today’s campuses. Adapting safety methods to decrease real anticipated safety threats is always necessary. Surveying available campus safety techniques will help craft integrated safety solutions and secure campuses from threatening and offensive activities. In this paper, we aim to survey the current literature on campus safety and classify the available techniques concerning their methods, environments, implementations, goals, and other factors. We also associate each mechanism with safety concerns that it addresses.

Phishing is one of the most common threats to students’ learning environments and academic institutes. Phishing attacks are a common type of social engineering attack used for fraudulent activities. In this type of attack, attackers aim to collect sensitive information from users in the scale of small or large range for the targets of malicious reasons. This research describes various mechanisms for phishing attack detection and prevention in higher education environments. Phishing attacks have different forms that can lead to devastating results, such as financial losses for organizations and individual users. Our research goal is to illustrate an adequate technique that can help reduce harm resulting from such attacks and increase information security awareness among users at all organizational levels.

2.1 Campus Safety Models and Methodology

Campus safety is an essential factor that helps to increase students' productivity. Students who feel safe and secure on campus are more likely to participate in collegiate activities and not worry about their being. Campus safety is necessary for Parents to feel comfortable sending their children to college.
Campus security protocols and safety continuously aim to secure students, employees, and other individuals’ well-being. Campus security management has been increasing in recent years with the national attention to campus safety, women’s safety, and personal safety. Available technologies are utilized to enhance student safety and security by identifying or deploying new methods. The safety threats that students and individuals face nowadays are broad, and many techniques are used for safety protection. Safety protection techniques can include monitoring software, wearable sensors, GPS trajectory, FRID-based approaches, and monitoring systems based on deep learning [21].

Campus security depends mainly on security personnel which is an effective technique to a certain limit. That can depend on the security personnel's ability to face change and recognize the dangerous behavior of students or individuals in the first place. Thus, the traditional supervision of campus safety that depends on manpower is not enough and is prone to safety risks. Developing a safety monitoring system can help with dangerous identification and speed detection on campus [22][23].

In the current trends, wearable Sensors are used as practical tools for tracking location and health issues using smartwatches and GPS tracking devices. Parents can remotely monitor the college student or younger children and communicate with college security. The main problem with these devices is the time gap. To avoid these problems, teachers would receive a message for students facing severe trouble on the school campus. Still, this technique is not reliable for children’s security because of the individual teacher’s attitude [24] [25].

Campus environments are associated with using practical mobility vehicles such as bicycles, tricycles, and scooters. Riding intelligent scooters is also vulnerable to road users and severe injuries during road accidents. Using the intelligent android system STEADI for monitoring the smart scooter rides. The implemented system uses a Wearable Gait Lab, a wearable underfoot force-sensing intelligent unit, as one of the main components. This system aims to help new students with no experience in riding scooters on campus avoid accidents and injuries by alerting the rider about sudden and unforeseen conditions [26].
Safety and security challenges and risks that are associated with student housing are one of the significant safety challenges students face nowadays. Facilities Housing is a significant physical infrastructure for both living and academic activities. The student housing facilities are used for student accommodation while running their academic program and can be located either on or off campus. Universities are responsible for student housing facilities located on campus, while private owners run off-campus housing facilities. It is significant [27]. It is helpful for campus safety to detect and analyze students’ behaviors on campus using CPS trajectory. Significant daily behaviors, their priorities, and abnormal behaviors using a graph-based approach [28].

In this paper, we aim to study the existing literature and the available mechanisms for campus safety. These techniques target specific locations (e.g., public spaces, residence halls, and campus neighborhoods). They also address specific environments such as colleges, schools, homes, student housing, transportation, etc. Several methods are provided in the existing literature focusing on monitoring systems, wearable devices, GPS trajectory, the use of mobile applications, use of machine learning. These mechanisms also have various goals and address different safety concerns. We survey these techniques and classify them concerning their methods, environments, implementations, and safety concerns. We also discuss these techniques and analyze their impact on different safety concerns. Our work will help set the stage for future research on campus safety. We will illustrate the weaknesses and limitations of existing methods and specify the required future work.

The remaining part of this paper is structured as follows. Section 1 provides a literature review of the existing techniques. Section 2 discusses and identifies the current limitations and weaknesses of the available methods. Section 3 provides our conclusion and future work.

2.2 Literature Review of Existing Campus Safety Mechanisms

In our survey, we classify existing research work based on four aspects: 1) campus safety method, 2) campus environment or location, 3) campus safety method goal, and 4) addressed safety concerns. Figure 1 displays the main classes of our taxonomy. In the campus safety methods, we will describe the main techniques used to achieve campus safety, e.g., wearable devices [22][26], and monitoring systems
The environment class includes the target location by each campus safety technique, e.g., student housing [28], and nighttime [30]. The third class of our taxonomy is the goal of the research work. Example research goals include safety assessment [26][27], and detection of criminal activities [31][32]. Finally, the safety concern of the research method includes the primary entity of protection or the type of offense. E.g., protecting human life [28], and fire safety [27]. The main subclasses of our taxonomy are as follows.

- Campus safety method
- Target environment
- Purpose/goal
- Safety concern

We classify the campus safety technique based on the utilized approach into campus devices [22][33], systems [34][31], and implementation methods [36][37]. We further classify work environments into location-based environments (e.g., transportation [26][28], student housing [27][38]) and specific environments (e.g., women’s safety [36], Nighttime safety [30]). Our subclasses for research goals include assessment [27], crime detection and prevention [32], and other purposes [27]. Finally, Safety concerns describe the type of crime or offense the research addresses.

<table>
<thead>
<tr>
<th>Paper</th>
<th>Safety Method</th>
<th>Purpose/Goal</th>
<th>Safety concern</th>
<th>Target Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>[22]</td>
<td>Wearable sensors</td>
<td>Safety enhancement</td>
<td>Human safety</td>
<td>Campus area</td>
</tr>
<tr>
<td>[29]</td>
<td>Implementation/Design</td>
<td>Safety monitoring</td>
<td>Human safety</td>
<td>Campus area</td>
</tr>
<tr>
<td>[26]</td>
<td>wearable things</td>
<td>Safety assessment</td>
<td>Safety conditions</td>
<td>Transportation</td>
</tr>
<tr>
<td>[27]</td>
<td>Safety Education</td>
<td>Safety analysis</td>
<td>Campus Crimes</td>
<td>Student Housing</td>
</tr>
<tr>
<td>[28]</td>
<td>GPS trajectory</td>
<td>Safety analysis</td>
<td>Crowd management</td>
<td>Campus area</td>
</tr>
<tr>
<td>[39]</td>
<td>Implementation/Design</td>
<td>Safety monitoring</td>
<td>Campus Crimes</td>
<td>Houses/Dorms</td>
</tr>
<tr>
<td>[38]</td>
<td>RFID-based tracking</td>
<td>Safety enhancement</td>
<td>Crowd management</td>
<td>Transportation</td>
</tr>
<tr>
<td>[36]</td>
<td>Wearable Device</td>
<td>Safety enhancement</td>
<td>Human safety</td>
<td>Woman safety</td>
</tr>
<tr>
<td>[40]</td>
<td>Deep Learning</td>
<td>Safety monitoring</td>
<td>Human safety</td>
<td>Campus area</td>
</tr>
<tr>
<td>[33]</td>
<td>Implementation/Design</td>
<td>Safety enhancement</td>
<td>Human safety</td>
<td>Campus area</td>
</tr>
<tr>
<td>[34]</td>
<td>Implementation/Design</td>
<td>Safety enhancement</td>
<td>Human safety</td>
<td>Campus area</td>
</tr>
<tr>
<td>[31]</td>
<td>Implementation/Design</td>
<td>Safety planning</td>
<td>Campus Crimes</td>
<td>Campus area</td>
</tr>
<tr>
<td>[32]</td>
<td>Mobile safety app</td>
<td>Crime Reporting</td>
<td>Safety conditions</td>
<td>Campus area</td>
</tr>
<tr>
<td>[37]</td>
<td>Mobile safety app</td>
<td>Safety enhancement</td>
<td>Safety conditions</td>
<td>Campus area</td>
</tr>
</tbody>
</table>
In the following subsections, I will describe the main techniques under each main class of our taxonomy.

![Sankey Diagram](image)

*Figure 13- Relationships of main safety goals, concerns, methods, and target environments of campus safety research work.*

The Sankey Diagram in Figure 13 describes the relationships among the existing campus safety research work concerning their research method, research goal, safety concern, and campus safety environment. The diagram shows such a relationship as data flow between parallel coordinates. For example, the diagram shows that most of the research focused on designing and implementing safe and secure campuses, targeting three goals: safety monitoring, planning, and enhancement. Another example shows safety assessment work on the lower part of the goal axis (second vertical bar) targets either general safety conditions or campus streetlights on the safety concern axis.
2.2.1 Classification of Existing Campus Safety Models

Safety management depends on various factors and techniques to ensure the environment's safety, including community safety, campus safety, housing facilities, and women’s safety [42]. There are various methods used that contribute to increasing the security of individuals. These methods include wearable sensors, GPS Trajectory, campus safety mobile Apps, deep learning, and monitoring systems. The following are some details about a summary of the methods used.

![Figure 14- Number of techniques utilized in campus safety.](image)

Figure 14- Number of techniques utilized in campus safety. divides the various techniques utilized for campus and environment safety are divided into seventh categories.

2.2.2 Campus Safety Methods

Emphasis on the safety of the college campus network since it is an essential part of campus life. Software monitoring system methodology is used to ensure the safety, reliability, and steady running of the campus network. The network monitoring system mainly provides data for network management and maintains a complete function of network management which is a vital tool for network maintenance [29]. The primary goal of network monitoring is to test the network security using a standard method,
safety inspection, to build a network in the network monitor to test the information in front of the web server.

2.2.2.1 **Campus Safety Devices**

The Internet of Things (IoT) has a vital role when it comes to campus and individuals’ safety and security. As many researchers highlighted the effectiveness of utilizing IoT for increasing safety and security in different ways, the gorgeous advantages of the IoT are the wiles connection which allows people to be monitored from a distance. In our campus safety and security research, we discuss the effectiveness of applied techniques of wearable sensor technologies and radio frequency identification (RFID) as an effective wireless communication method to enhance campus safety and security.

2.2.2.1.1 Wearable Sensor Technologies

The advanced innovations include sensors of different sizes, types, and functionality to help the environment. Even though it was traveled, wearable sensors play a vital role in collecting data. Wearable sensors are used widely to monitor human activities related to safety, security, medical sciences, and sports. Wearable sensors can detect abnormal activity and unforeseen situations that are associated with people’s safety or health conditions. Wearable sensors and Gait Lab sensors are different mechanisms that are used for people’s safety. The difference is that wearable sensors are used for people’s safety and depend on wireless node structure when wireless nodes are connected in the form of bidirectional connectivity when the identity of one entity can be taken by another or each other. Wireless connectivity can increase the speed of the message transfer, which can help the security services take any required action to ensure the safety of the people associated with two nodes in case of unnecessary movement [22], [41].

The central processing units are connected to all nodes and will be responsible for tracking the movements of all nodes on schools or college campuses, and can focus on staff members, teachers, and students to ensure safety and productivity [26], While Wearable Gait Lab sensor is used for safety protection that can monitor the smart scooter riders. The Gait lab sensor can measure ground reaction
when worn under the person’s foot [36]. The advancement of enhanced technology includes smart wearable devices for safety. The device will alert and automate the emergency system by using three different sensors: a pressure sensor, plus rate sensor, and a temperature sensor.

Focus on wearable sensors. These systems have different functions, such as bio-signal examining, signal analysis, and biofeedback. They consist of sensors, data aggregators, and data processing units connected by wired or wireless communication modules. The main goal is to develop and introduce a system that increases performance accuracy, prolongs battery life, and reduces cost. Lately, Various personal computers and phone devices are Bluetooth smart ready, which is simply the operation of wearable healthcare system technologies with a focus on Bluetooth low energy technology [25].

2.2.2.1.2 Radio Frequency Identification (RFID)

RFID Radio frequency identification (RFID) consists of two wireless system components: tags and readers. The reader is a service with one or more antennas that emit radio waves and receive signals from the RFID tag. Tags used to communicate can be passive or active; the reader powers passive RFID tags and does not have a battery, while active RFID tags are powered by batteries [44]. RFID includes different types of applications that use many different frequencies to transfer data. The available FIRD application includes personnel tracking, monitoring patients, which provides medication scanning and out-of-bed detection, and inventory monitoring.

RFID, Bluetooth, and Wearable Sensors system is a tag system that can support the school management system by verifying students’ attendance automatically. The RFID tag can provide unique identification information for each student. When students cross the RFID scanner device, it will get the student’s unique identification information and then send it to the server system. RFID server software can access this information to track the student that belongs to the presented information [22][26][36].

2.2.2.2 Campus Safety Systems

GPS Trajectory, Campus safety mobile app, and monitoring system are three major components of campus safety. GPS Trajectory devices can be integrated into campus safety systems and utilized to
track and monitor students, staff, faculty, and other members of the campus community. Campus Safety Mobile App can provide a large range of features for individuals, especially students, and enhance their safety such as reporting mechanisms. The Campus Safety Monitoring system uses a dictation system based on deep learning behavior. In the following section, we explain the current literature review and the contribution of the campus safety system to maintaining a safe campus environment.

2.2.2.2.1 GPS Trajectory Technology

Li et al. [28], this research methodology focuses on a graph-based approach that uses GPS trajectory to detect students’ mobility patterns. This analysis used five steps: data preprocessing, dwell point extraction, behavior graph creation, priorities, and abnormal analysis of students’ campus behavior.

Step 1, data preprocessing: to better visualize trajectory and analysis results, the GPS coordinate is converted to Baidu coordinate, and the GPS drift point caused by noise is removed [28].

Step 2, dwell point extraction: is used to reduce inherent inaccuracy in GPS measurements caused by noise and other reasons; density-based spatial clustering of applications with noise DBSCAN is used to extract dwell point expressed as (B, L, t) where B is latitude, L is longitude and t is the time students stay.

Step 3, behavior -Graph creation includes two measures and advantages. Firstly, it focuses on topological relations between behaviors instead of geometric simplification. Secondly, it partitions all trajectories by treating them as a complex network instead of partitioning individual trajectories separately.

Step 4, priorities, and abnormal behaviors analysis: in this step, the K-core algorithm is utilized to recognize students’ priorities behavior since student behavior has different priorities. K-core is often used to analyze different essential clusters in the network. Abnormal behaviors are often independent of normal behaviors, and a closeness center degree is often used.

2.2.2.2.2 Campus safety mobile applications

Mohammed et al. [37] This research mainly focuses on mobile applications and the KMSAFE APP, a campus safety App. KMSAFE application allows users like students, staff, and faculty to send a message notification to the campus security department using a small Bluetooth button to their
smartphone. The button can be kept in a critical chain or personal bag; they need to push it for three seconds. In this case, campus security will receive real-time emergency notifications from the victims in danger. This message includes the victim’s personal information, map position, and surrounding area images.

Et al. Kumar [30] Street safety using mobile applications since mobile applications become a core part of a user’s life. Walking during nighttime can be risky at some point. This mobile research system helps users know which street is lit at night by providing approximate locations of streetlights around a specific area to engage women and children to step out during nighttime without concern. The method’s central part is a mobile application that views the location of nearby streetlights in Map View. Map View is a part of the Google Maps API, which displays a map. When the application is opened, the user’s mobile sends a POTS request to the server. It then takes the user’s (coordinates) from the request body and queries all the nearest streetlights. The locations of the streetlights are then sent back to the user. The location of the streetlights is sent back to the user, and the data will be transferred to the server where the processing takes place.

2.2.2.2.3 Campus Safety Monitoring

The campus safety monitoring research focuses on safety hazards utilizing a dictation system based on deep learning behavior. The system integrates hazardous behavior identification, license plate tracking, and vehicle speed management [24] [40]. Other methods focus on the campus network since it is an essential part of ensuring the safety, reliability, and steady running of the campus network. The primary goal of network monitoring is to test network security. The most common myth used for network inspection is to create a network in the network monitor to test the information in front of the web server. That will help detect suspicious attacks and illegal access to the system. The security system can be wireless or non-wireless networks [29] [39].
2.2.2.3  Campus Safety Reporting

The primary goal of the applied methodology, a campus crime report system, is to provide an online system that provides an anonymous and secure mechanism for both victims and witnesses to report crimes to the police. The system is being implemented and tested on a university campus, and the potential users, which include students, staff, and faculty, were surveyed to determine their intent to use the system [32]. Respondents found the online system accessible, helpful, and safe to report a crime, but the type of crime and the urgency of response is a contributing factor in the determination to use the system vs. reporting it to a live person [32].

The survey questions were grouped into four sections. The first section asks participants general information questions such as title, department, age, confidence, trust in using the Internet to conduct transactions, and their trust in the Department of Campus Safety’s commitment to resolving on-campus crime. Section 2 focuses on asking participants about their probability of reporting crime-related information. Section 3 asks participants about their attitude toward using both voicemail and an online system to provide crime information. Finally, section 4 asks different questions regarding the effectiveness and efficacy of voicemail vs. online systems to report crimes.

2.2.2.4  Campus Safety Model Design and Implementation

The security of the campus network is an essential part of campus safety, as well as the security and management of personal computers. Implementing a campus network monitoring software system on the campus network security management can help the campus to prevent and detect malicious attacks from unseen network security threats, thus avoiding the university network from various types of cyber-attacks. The network monitoring system is the basis of network supervisory control. Through the network, it can analyze the information on the network, which can help to find hacker’s invasion in time and stop any unsecured activities and help assist the administrator for the network security management and maintenance in ensuring the normal operation of the university network.
The main goal of the network monitoring system is to test the campus network security. The most common method used in safety inspection is building a network monitor network to test the information in front of the web server. It can detect suspicious attacks and any illegal access to the system, then issue instructions to block information containing some patterns through the network filter, thus maintaining the network server and operation stability. The safety monitoring system technology through hardware and software to achieve real-time inspection, data flow, and system intrusion signature database of data comparison [29].

2.2.2.4.1 Deep Learning Methods for campus safety Model.

Campus safety monitoring system based on deep learning methodology. The utilized system is divided into three different sections: recognize behavior, license plate tracking, and vehicle speed measurement. The vehicle speed detection algorithm is based on a virtual coil and is mainly divided into several processes that include traffic video using the MATLAB platform and a set virtual coil that uses the manually set virtual coil. Because this coil is viable and can easily be used in the first frame image, and calculation of the gray difference between virtual coils VLI and VL2[24].

2.2.2.4.2 IoT Technology for Campus Safety Models

The Internet of Things introduced several solutions for campus safety and security with the ability to connect different security technologies, which provide real-time alerts and data-driven insights. The IoT is classified as a network of physical devices with unique IP addresses for internet connectivity and communications. In general, IoT devices are intelligent devices that combine capabilities and enable advanced automation functions that allow for streamlined security protocols. IoT devices collect data, and unlike other traditional security equipment, they have the power and ability to exchange and analyze the data they collect. The utilization of IoT can guarantee that data can be processed and relayed much faster than old-fashioned systems [43].
IoT is used to protect women and children from significant environmental threats [22][36]. In [22], the applied techniques utilize intelligent sensors to increase children’s safety on the school ground. The smart sensor consists of two nodes connected by Wi-Fi. Data is sent to the server for analysis and processing to identify any abnormal activities for all parts and increase campus safety [22]. While in [36] a wearable system for women is introduced, the system consists of three sensors: temperature, pressure, and pulse rate.

The temperature sensor is used to detect any deviation in the temperature, the pressure sensor is used to detect if any pressure is being applied to the woman beyond an acceptable limit, and the pulse-rate sensor is used to detect abnormalities in the woman's pulse rate. The combination of these three-sensor readings is used to detect any critical situation. The wearable sensor system device also provides a push-button for women to press in case of feeling unsafe. The buzzer is activated to alert friends and family around her that the woman is in a dangerous situation. Then the woman’s location in danger is detected using a GPS module. GSM is used to send messages to relatives.

2.2.3 Campus Safety Environments, Goals, and Concerns

This section describes the main environments addressed in existing campus safety research. We also discuss the main safety goals and concerns with current research.

2.2.3.1 Campus Safety Model Environments.

The primary goals of this research are to increase safety management and to ensure a better environment for everyone, thus increasing the productivity of individuals and students. The measurement of campus and environment is divided into two sections: location-based environment and specific environment. The location-based environment, [38][28] includes public space, student housing facilities, and campus neighborhood [21][37]. The specific environment includes women’s safety and nighttime in the scope of increasing city safety [36][30].
The main goal of increasing safety in public spaces is to ensure that students and individuals are aware of the environment and move around the city reasonably and without fear. In [26], the smart scooter assistance system is consistent with four modules designed to increase the smart scooter tide around the campus environment since many students will rely on the smart scooter as one of the campus’s available transports used among students [28].

2.2.3.2 **Campus Safety Model Goals.**

The primary goals of campus safety are to provide a safe, secure environment for students, staff, and faculty. [45] Security and security are significant factors for the university to accomplish its mission and introduce a productive and safe environment. The applied techniques to reach the campus goal are maintaining a productive and safe environment for everyone, including assessment of risks, detection of criminal activities, and safety and security risks [27] [28]. In [28], using a graph-based approach, the GPs trajectory is used to detect and analyze students’ primary daily behaviors, including priorities and abnormal behaviors. There are also other safety goals include, which include women’s safety, especially during the nighttime, and children’s safety on the school grounds. The main goal is to create a safe and productive environment for campus students and the community to improve productivity and high performance; People are more likely to be productive in a safe and secure environment.

2.2.3.2.1 **Campus Safety Model Assessment**

The risk assessment, management, and measurement of threats are applied techniques to Virginia Polytechnic Institute and State University. Risk assessment is used to reduce any type of risk around the campus. The process of identifying hazards that could affect the campus can include hazard identification which includes potential risks, hazard events, or liabilities. Hazard Characterization: the evaluation of which personnel, property, income, or assets are most vulnerable to injury or damage from these hazards by severity and frequency.
Exposure Assessment: estimation of potential losses and risk characterization: the prioritization of various risk exposures. Risk management includes the applied policies, procedures, and practices associated with identifying, analyzing, and assessing risk exposure and appropriate strategies to reduce and minimize unacceptable risks. The teats assessment includes a fact-based investigative analysis approach that evaluates an individual’s behavior that poses a risk to their safety or the safety of others [46].

2.2.3.2.2 Detecting Criminal Activities

This research aims to detect crime in video streams such as CCTV cameras. In addition, a module for detecting faces in these streams is also done using a previously implemented method known as Triplet Loss. This research proposed two modules for showcasing the given problem. The first method is taking CCTV steam and recognizing the faces, and the next model discusses tackling crime detection using deep learning techniques. The face recognition module uses Open-CV’s built-in DNN Deep Neural Networks module to train a model primarily built to differentiate faces and recognize them. It is done by identifying faces through given data, performing data preprocessing, and then proceeding with training on embeddings using the Triplet Loss function.

After these embeddings are calculated, the model is then used to recognize such faces. This can be done simply by loading the model and using a webcam to distinguish the face by surrounding it with a bounding box and giving a secret parameter. The Crime detection technique initially selects the input videos that process criminal activity and videos with standard or no criminal activity. After performing the main processing steps, which include data augmentation and converting the videos to image frames, it is trained through a pre-excising Resent (artificial neural network) architecture through which accuracy and other evaluation metrics are observed. The trained model can detect crime by playing virtual webcam streams through a mobile phone; overall, this research suggests an end-to-end pipeline to tie both modules and use them simultaneously to identify any crime or abnormal activities on the campus ground [21].
2.2.3.3 Campus Safety Model Concerns

Campus safety and security are essential aspects of college education. The Department of Education is committed to assisting schools nationwide in making a safe, natural environment to learn and keep learners, parents, and workers highly informed about campus security. The Crime Awareness and Campus Security Act of 1990 is committed to ensuring universities and high education institutions are in full compliance with the act. Enforcement of that act is the priority of the Department. Various topics are a concern and covered by the U.S. Department of Education to increase awareness among students, staff, and employees and increase campus safety and security. The safety and security topics related to safety measurements include human life threats, hate crime, fire, weapons on campus, drugs, cyberbullying, identity theft, and online security in college and liquor [47][48].

2.2.3.4 Campus Safety Model Classification

There are various methodologies introduced to improve campus safety and security. Several introduce methods that mainly focus on student safety in the campus ground around the city. In contrast, others focus on the security system network and security system of the university. It is essential for students, staff, administration, and others to feel secure in the campus environment. Securing some risk factors does not mean students or everyone on the campus is fully secured. Therefore, when people think about safety and security, all the factors and concerns of a system need to be included to develop a security system that grants a secure campus and safe students.

The safety and security of the educational institute are essential to support a healthy and productive environment for students, workers, and individuals. There are various techniques used to increase and support the environment. The methods used depend on the advancement of information technology to support a safe campus environment which devices, systems, and implements. The device methods focused on wearable and FRID-based sensors techniques. The system methods focused on the GPS trajectory, campus safety mobile App, campus crime report system, and monitoring system.
The implemented methods include design and implementation, deep learning methods, and the Internet of Things. These methods applied in different environments include location-based environments and specific environments. The location-based environment includes public space, student housing facilities, campus neighborhoods, and transportation. Specific environments consider women’s safety and nighttime safety around the city. The applied methods are various, but in the end, all methods have a common goal: keeping a safe and secure campus environment for students, faculty, and staff. In addition, increasing the productivity of individuals can lead to long sustainability of the educational institute that works to maintain a healthy learning environment for everyone.

2.3 Campus Phishing Attack Detection and Protection Methods

According to [49] the internet has been an important part of human lives. As of October 2020, 59% of the global population are active users of the Internet. This number significantly increased due to the Covid-19 pandemic, which led to a corresponding increase in the cyber-attack rate. Verizon Data Breach investigations in 2020 reported 67% of data breaches for social engineering attacks like phishing or email phishing attacks. There are different types of phishing, including SMS phishing, vishing or VoIP, deceptive phishing, spear phishing and whaling, link manipulations, in-session phishing, etc. This chapter section focuses on recent detection techniques as well as protecting students of higher institutions from malicious attacks.

To focus on recent detection techniques as well as protect students of higher institutions from malicious attacks, we selected papers between 2018 and 2023, that focused on students and faculty awareness and explored a wide range of methodologies in achieving this goal. The study contains a review of 12 papers that have been published in IEEE.

Ripa et al. [49] identified phishing as the number one cyber-attack attack vector. The authors describe three main popular ways of phishing attacks which include attacks by phishing URLs, phishing websites, and phishing emails. In addition, due to the increase in popularity of social media and online gaming, phishing attacks have now also been extended to these forms of media. In this paper, the authors
presented machine learning models to detect the three main types of phishing attacks, a Twitter spear bot was built using machine learning, and experiments were carried out for each phishing attack type.

- For phishing URL detection, various classifiers were used with higher accuracy and focused on the timing to train the dataset. In an implementation, a phishing URL dataset was selected, and the XGBoost algorithm was used to detect the phishing URLs. XGBoost was particularly chosen because of its high accuracy and efficiency.

- The Naive Bayes (NB) Classifier was used for Phishing Email detection to differentiate between ham (good emails) and spam emails. The NB classifier separated the emails by words like text categorizing.

- For Phishing website Detection: Data was collected from the UCM Machine Learning Repository, a collection of databases and generators of datasets. The data was split into two parts; where the 70% ratio was for training purposes while the remaining 30% was used as a testing dataset. After these, five models were used to train the dataset separately, these models are Random Forest Classifier, Decision Tree Algorithm SVC, KNN, and Logistic Regression.

The results of the phishing URL detection showed that the XGBoost classifier gave an accuracy of 94.44% with less time. In comparison, the phishing email detection with Naive Bayes Classifier had an accuracy of 95.15%. Finally, of all the models applied for phishing website detection, the Random Forest Classifier had the highest accuracy to a degree of 96.80%. Having three models to cater to each of these scenarios might be challenging whilst implementing them on an application. However, a single model can be trained to cater to all three types of attacks, with features that would produce high accuracy in a limited time.

Athulya and Praveen [50] discussed various phishing attacks, the latest phishing evasion techniques, and approaches to detect these attacks. The authors divided phishing detection techniques into 3, namely, user education, website content-based, and website non-content-based detection.

A. User Education technique: Athulya and Praveen emphasized educating users via social media and emails to ensure that they can recognize phishing attempts on themselves by themselves.
B. Website Content-based technique: This technique uses the features of a web page to detect phishing web pages by extracting its features. The authors further divided this technique into 3 approaches of detection which are as follows.

- **Visual Similarity-based phishing detection:** With this technique, phishing websites are identified by the presence of images, Flash, and Java Applets that are inserted into the website instead of HTML text, using a signature/similarity ratio. This signature/similarity ratio is often generated by taking similar features of a website and stored into a database.

- **Machine Learning Approach:** The authors further suggest that the ML algorithms be applied to the previously derived features to detect phishing attacks. Some of these algorithms, including Random Forest, Sequential minimum optimization, Logistic Regression, Support Vector Machine, etc. are applied based on their efficiency. To a large extent, the efficiency of this approach usually depends on the size of the dataset, selected features, and ML classifiers.

- **Heuristic-based Approach:** This technique extracts common properties such as Hyperlinks and URLs from fake websites that exist to find new phishing websites.

C. Website Non-Content Based technique: This technique is further divided into three which are as follows

- **List-Based Approach:** The main purpose of this technique is to check the status of a web page if it is legitimate or phishing. This approach can either be blacklistedor whitelisted. The Blacklist approach stores phishing URLs, DOM tags, SPAM URLs, and other information. It is generally set as a browser plugin; examples of tools implemented with this approach are Google’s safe browsing (GSB), Opera, etc. On the other hand, the Whitelist approach, the opposite of the blacklist approach, uses a list of legitimate styles, URLs, DOMs, and digital certificates for comparing fake websites.

- **Search Engine-Based Techniques:** In this technique, identity information is retrieved from websites and search engines to determine the authenticity of the queried websites. This technique
can detect a zero-hour phishing attack and has a high response time.

- **Crowdsourcing:** This detection technique relies on the rating given by users to determine the legitimacy of a website. An example of a crowdsourcing tool is the Web of Trust (WOT) browser extension, a program that monitors user behaviors, validates ranking, etc. In practice, a traffic green light appears when a user rates a legitimate website, and red color alerts indicate potential threats, while yellow colors mean warnings while using the website.

The detection accuracy of each of these techniques can be seen in the table below, ranging from low level to high level of accuracy.

<table>
<thead>
<tr>
<th>Anti-Phishing Solution</th>
<th>Zero-Hour Attack</th>
<th>Response Time</th>
<th>Detection accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>List Based Approach</td>
<td>No</td>
<td>High</td>
<td>High</td>
</tr>
<tr>
<td>Heuristic-based &amp; ML approach</td>
<td>Yes</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Visual similarity approach</td>
<td>No</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Search Engine Based</td>
<td>No</td>
<td>High</td>
<td>Medium</td>
</tr>
<tr>
<td>Crowdsourcing</td>
<td>No</td>
<td>High</td>
<td>Medium</td>
</tr>
</tbody>
</table>

As much as these detection techniques are very important and useful, the authors did not indicate what tools or standards were used to determine the accuracy of the technique.

In the paper [51] the authors identified phishing, crypto-jacking, and ransomware as recent threats to cyber security. They further described the specifics of phishing emails and provided advice on protection against these attacks. Some of the phishing attacks mentioned by the authors include sending phishing emails with attached files, malicious pdf files, cloud services, banking trojans (Dyre, Riltok), phishing in the cryptocurrency world, and Cross-Site Request Forgery (CRSF) attacks. To mitigate these attacks, the following methods were presented to protect oneself from phishing attacks.

- The use of two-factor authentication such as hardware tokens for unique password generation for a session.
- Constant check of sender addresses; users are advised to always take a close look at the sender’s email address before opening links or attachments sent in an email.
- Updating software regularly on one’s computer as well as using licensed antivirus software.
- Regular backup of important files
- Users should never reply to strange emails.
- Users should never call phone numbers on suspicious letters or emails.

In addition to the prevention techniques listed above, regular training should be held to keep users informed on new tactics used by attackers as well as remind them of measures that can be taken to protect themselves.

Younis and Mohamed [52] developed a framework for protecting against phishing attacks. They identified a lack of user knowledge as one of the main reasons phishing attacks have many victims. The authors particularly focused on creating a mitigation solution for Arabic users as they perceived that not too many solutions been created for them, even though phishing attacks may have a more severe effect on them. The developed framework was divided into two parts, namely.

A. The training Component: This component consists of the user training model and the level of assessment.
   - The User Training Model: The goal of this model is to improve the users' awareness by providing efficient training, after which the users are tested at the end of the training to determine their level of awareness of various types of phishing, either emails, websites, or SMS.
   - Level of Assessment: Here, the users are being tested on real examples of phishing attacks, and based on their results, they are classified as beginner, intermediate, and advanced users. Users who pass the exam at the beginner level are tested to take the next level assessment test (Intermediate). When users pass these tests, they usually proceed to the next level; otherwise, they retake the exams of the same level till it is ascertained that they fully understand what is expected of them at that level.

B. The Gamification Component: This part of the framework s two databases: the Arabic dictionary and the phishing/legitimate dataset. The databases were used to build games based on challenges and assess the knowledge delivered by the training system.
The developed framework is aimed to train Arabic users in detecting phishing attacks in the Arabic language and assess delivered knowledge using gamification. The user training aimed to train users to reach the right level of awareness and expertise in detecting these attacks, while the gamification parts the users to real test to detect real and fake phishing attacks using a challenge-based games approach. In our opinion, phishing attacks have the same common goal in all cultures and languages; therefore, we believe that the authors can develop the framework for a wider scope audience.

In this paper [53] Ajay U. Surwade et al. described the development of a filter-based origin-based information email called an origin-based filter which blocks these phishing emails at the MTA or MDA level by extracting header part information of these emails using the blacklist approach. The OBF is tested using standard dataset public phishing crops, and the result shows 100 percent accuracy of this Origin Based Filter (OBF).

In this method, the known phishing and spam emails are used to extract the important features of an email that can include from, subject, to, date, return-path Delivered, X-Original, etc. These features are kept as the header of Spam e-mails plus a header of phishing emails. In this case when new email research to MTA or MDA the same feature is extracted from the email. After that, the features are Compared with the header of Spam and the header of a phishing email, and the email is tested for phishing or non-phishing email. Public phishing corpus datasets are downloaded for features extraction, and a Python program is used for extracting information from emails using the Python library RFC822.

In this case, the program has extracted all sender’s email addresses, which are named Blacklist addresses because the public phishing corpus is a standard dataset of phishing emails. On the other hand, another Python program is used to test new phishing emails, and the program accepts new emails or testing. The program extracts the field information as an example ‘To’, ‘Return path’, and ‘form’. The result of this research shows that the proposed OBF gives 100 percent accuracy of classification, even though whitelists and blacklists are not sufficient to block phishing emails, when combined with other filters such as OBF it would certainly reduce the load on the complete phishing classifier.
In addition to the proposed machine learning algorithms by[49][54], Sushma et al. [55] also explored the use of machine learning, in this case, deep learning to detect phishing websites by building a system that takes in the URL to be verified, and read the host/page based features and perform lexical feature extraction on the characters of the URL, after which these features are evaluated using machine learning methods, a classifier is selected and implemented, the authors selected Support Vector Machine and Random Forest Classifier. On completion, the analysis of the performance of the classifiers was carried out to determine the best-performed classifier. The results showed that the Random, Forest Classifier achieves a higher accuracy level than the Support Vector Machine.

Gajera et al. [54] developed a novel approach to detect phishing attacks using Artificial Neural Networks combined with pharming detection. The authors trained URL features of phishing and legitimate websites on Artificial Neural Network (ANN) for accurate classification, they also implemented provisions to detect pharming which is simply DNS poisoning in which a user is redirected to fake websites by changing the IP address at the DNS server. To achieve this, the block system contained a first phase where the authors used a feature engineering approach to extract important variables. Then, the Artificial Neural Network was used to build a model and the model predicted values between 0 and 1 to predict the legitimacy of the URL. In the second phase, they performed steps to detect pharming. The developed system was divided into three parts which include:

- **Feature Engineering:** Here, the authors identified 15 features that could be used to analyze URLs, such as IP address, URL length, number of links pointing to the page, presence of @ symbol, number of subdomains, domain registration age, URL redirect, etc.

- **Classification Algorithm:** A classifier was built and trained with the identified features using a feedforward Neural Network with 3 hidden layers and one input and output layer. Neural Network was chosen because of its ability to classify URLs effectively as well as its high level of accuracy.

- **Pharming Detection Technique:** The authors detected pharming by querying a local and a global DNS to get IP addresses and check for matches. When matches are equal, no pharming is detected. However, if they are not equal, they proceed with web page comparison by visiting the IP addresses and taking
screenshots to check whether they are the same. When a website turns out phishing-infected in the first phase, the website is declared a phishing website, and the second phase is skipped. If it turns out to be legit in phase one, then phase two begins.

For the experimental setup, 16,053 datasets of phishing websites and 7,974 legitimate websites were collected from PhishTank and fcsit. Animas repositories, respectively. After this, feature engineering was performed on the dataset to extract the features and later stored as a training-compatible dataset .csv file. Finally, the dataset was split into a train-test split and used in TensorFlow. The results showed that the detection system had an accuracy of 98.77% and a precision of 97.56% using a confusion matrix system. The authors concluded that the developed system could detect pharming and phishing quite accurately. Although the authors intend to make the developed system into a plug-in for google chrome that can be used for all users, we also think that it can be a background application/service installed on users’ systems to detect these URLs on other applications used by users such as Skype, Slack, Teams chat, etc.

In a paper [56], Wang et al. developed a cost-effective OCR implementation to prevent phishing on mobile platforms. The authors applied a novel method called the OCR (Optical Character Recognition) in a mobile platform and further implemented a prototype to determine the applicability of the novel method. The author’s method required Google Optical Character Recognition (OCR), Google Search API, and also installation of a MITM proxy (man-in-the-middle proxy) on the test server. The implementation contained four phases which are on the test server. The implementation contained four phases which are.

- **Interrupt Traffic and Redirect to Server:** The authors wrote a python named ‘mitm.py’ in mitmproxy. The ‘response’ of the function was used to handle the traffic flow. In this proxy, firstly, the URL address is retrieved from the request sent by the mobile. Then the proxy redirects the URL address to the browser vendor server to verify its security through the OCR function as the second step.
- **Implement OCR Functions in the test Server:** After the server receives the URL request, the OCR manipulation starts. This manipulation includes 4 steps which are.
• Parsing the URL and extracting all images
• Recognizing the image content
• Searching official websites’ URL
• Verifying the security of the parsed URL.

- Execute Response in mitmproxy: According to the received result, the proxy executes different responses if the URL is verified to be a malicious link, then the proxy would not redirect the URL request to a real site and send a warning page to the user directly. Otherwise, a redirect request to the real site is made and a parsed response is returned to the user.

A total of 60 URLs were collected from PhishTank, which satisfied some malicious URLs conditions specified by the authors, which were evaluated using a Safari browser on IOS alone and Safari with OCR implementation. The results showed that the Safari browser itself detected only 40% of malicious URLs while Safari with OCR detected up to 90% as malicious sites. The developed technique has a reasonable level of accuracy, and we believe it can be implemented at all levels of organizations. Below is an image of the results obtained during the evaluation between Safari and Safari with OCR solution.

In a paper [56], Wang et al. developed a cost-effective OCR implementation to prevent phishing on mobile platforms. The authors applied a novel method called the OCR (Optical Character Recognition) in a mobile platform and further implemented a prototype to determine the applicability of the novel method. The author’s method required Google Optical Character Recognition (OCR), Google Search API, and also installation of a MITM proxy (man-in-the-middle proxy).
In [57], the authors designed a novel tool capable of detecting phishing attacks and finding solutions to such threats. They also described a process that can be used for developing Scrum-based implementations for feature selection, automatic learning, and Neural Networks. Also, the tool developed can detect and mitigate phishing attacks registered in an e-mail server.

The applied methodology contained two parts, namely the design process and the development process. The design process was based on an architecture diagram that posed a view of the developed system. The authors presented the architecture of the system to allow browsing, detection, reading, alerting, and mitigation of phishing threats. In practice, the process such that when a new email is received on the server, the email is sent to the client and processed in the MatLab software. After which, the characteristics of the email are extracted using the feature selection algorithm.

Finally, the learning vector is generated using Neural Networks Algorithm; after selecting the characteristics to determine if an email is phishing or ham while in the development phase, the author implemented the proposed design. To test the accuracy of the proposed software, the authors experimented with a total of 3000 emails and classified the emails into three equivalent amounts of 1000 datasets each. Every classification was done based on the time, for example January to March, April to June, and July to September. The obtained results showed that the proposed software detected phishing emails to a percentage of 8.8%, while the blacklisted analysis detected 5.6%. Ona et al. concluded that the implemented algorithms complement each other during detection, which makes the software very
promising as it has an average accuracy of 93.9%. In our opinion, other machine learning algorithms can also be explored to improve the accuracy of the software in future versions.

Rose et al. [58] designed an intelligent phishing detection and prevention model using the Chrome extension. The proposed model was based on a machine-learning self-destruct algorithm. The rules employed in the model focused on URL-based web characteristics which have been identified as a feature attackers rely on to redirect victims to simulated sites. In an implementation, the authors extracted a total of 16 URL features such as IP address, URL length, presence of HTTPS in the domain name, the use of iFrame, presence of a status bar, anchor tags of href domains, degree of a subdomain, use of mailto, etc. Once a feature is extracted, it is encoded into values of -1 for legitimate, 0 for suspicious and 1 for phishing. After the feature extraction, three supervised machine learning algorithms (Support Vector Machines (SVM), Random Forests, and Artificial Neural Networks) were selected based on their performance on classification problems.

To evaluate the machine learning Algorithm, the authors obtained two categories of datasets; the phishing dataset from PhishTank and the ‘Phishing Website Dataset’ from the UCI Machine Learning repository while the non-phishing dataset was generated manually. They had a dataset of 11,055 URLs, of which 6157 instances were phishing and 4898 were legitimate. The three algorithms were evaluated based on the effectiveness metric values in terms of performance, after which the accuracy, specificity, and sensitivity were analyzed. To test the algorithm, a browser extension, Chrome, was embedded within the proposed algorithm. The results showed that Random Forest had the highest sensitivity with a value of 89% while SVM performed the best in terms of accuracy and specificity with values of 90.05% and 93% respectively. To improve the accuracy of the SVM classifier, more URL features can be added to train the model for URL extraction.

Altamash and Singh [59] in their paper Reconnaissance of credentials through phishing attacks & its detection using Machine Learning, simulated a phishing attack to steal user’s credentials without their knowledge by creating a clone of two major websites such that the pages appear real to the victims as the
actual websites. In addition, the authors presented ways of detecting phishing methods to prevent loss of data by using machine learning algorithms.

To simulate the phishing attack, the authors created a clone of Linked-In and Facebook login pages and then removed the sending parameters from the real address to the hacker’s address. On the server hosting the fake websites, they connected the pages to a database, such that after a user puts in their credentials which are the email and password, a file named ‘logshai’ is created on the server which contains the session_key and session_password. After this file is created, the user is then redirected to the original LinkedIn login page. Upon completion of the redirection, the user is left confused and made to believe that some internet problem occurred. The same process was also implemented for the Facebook simulation. Furthermore, the authors showcased a method of classifying malicious URLs from authentic source URLs by using these algorithms; AdaBoost, Kernel Approximation Classification and Multi-layer perceptron classifier, and stochastic gradient descent classifier.

The authors stated that one-way users can abstain from being a victim of a phishing attack is to consistently go straight to the website they intend to visit rather than clicking on hyperlinks. They also emphasized the importance of users paying attention to pages where they supply their information, as sharing information with dubious sites can lead to loss of money as well as tension among countries. In our opinion, the simulation of the phishing attacks was very convincing and provided good awareness. However, they did not present the accuracy or efficiency level of machine learning algorithms showcased for detecting malicious URLs.

Abedin et al. [60] proposed a software capable of detecting phishing attacks by differentiating between legitimate websites and phishing websites using machine learning classification techniques. They focused on using three machine learning algorithms: K Nearest Neighbors (KNN), random forest, and logistic regression.

To experiment, a dataset was obtained from Kaggle, a public repository that contains a large amount of dataset collection used for training machine learning models.
The obtained dataset had 32 attributes which are the URL features such as Abnormal URL, Using IP, HTTPS, Long URL, InfoEmail, Subdomains, Favicon, Website Traffic, etc. After obtaining the data, they performed feature scaling to normalize the independent variable of the dataset to a fixed range to handle variance in the value among different independent variables. The dataset was split into 2 parts, with 80% being the training dataset and 20% for testing.

The train test split was done using the Scikit-Learn library in Python. Each of the machine classifiers approached the problem differently, which is described below.

- **K- Nearest Neighbors (KNN):** This classifier used 80% of the labeled data to derive a function that predicts if a website is phishing or real.
- **Random Forest:** A supervised learning algorithm that uses a set of decision trees to build the forest.
- **Logistic Regression:** A statistical model that uses a logistic function to model binary dependent variables. Here, 80% of the labeled data derived a logistic function that predicted if a website is phishing or real.

The results of the experiments indicated that the Random Forest Classifier had a performance with a precision of 97%, an F1 score of 97%, and a recall of 99%. This showed that the proposed model is fast and efficient. In as much as the model's efficiency is to a good percentage, however, the model only uses the URL and ignores every other resource in its analysis; this could be a limitation as those other resources can be exploited for a phishing attack.

### 2.3.1 Results and Findings

We discovered that students are often left out in training organized by organizations for security awareness because more focus is placed on the faculty. The survey of these studies showed that many students cannot identify phishing emails if they ever receive them. Machine learning algorithms have a very high accuracy level in detecting phishing websites, and the most reliable way of identifying phishing websites is through feature engineering by extracting the features of the website and URL in question.
In this research, we focused on Phishing attacks targeted toward Campuses (students, staff, and faculty). For further research, we hope to extend this study to include vishing and smishing as fraudulent people are now employing these forms of attacks to retrieve information from unsuspecting users. In addition, the context of this future work would include different individuals across generations as well as organizations.

Phishing, a sort of social engineering, has been identified as one of the most common means of cyber-attack today, with its goal being collecting sensitive data such as passwords, credit card information, etc. from the user by creating clones of legit websites and directing users to these pages to steal their information.

2.4 Summary of Campus Safety Models and Protection Methods

Current research introduces numerous safety techniques to support an individual's safety during the day and night lives on campuses. Surveying available campus safety techniques will help craft integrated safety solutions and secure campuses from threatening and offensive activities. In this Chapter, we provide a survey of the current literature on campus safety and security. We classify the available techniques concerning their methods, environments, implementations, goals, and other factors. We also associate each mechanism with safety concerns that it addresses.

Phishing is one of the most common threats to students’ learning environments and academic institutes. This chapter also describes various phishing attack detection and prevention mechanisms in higher education environments. It also lists multiple ways of detecting phishing websites such as extracting information from the URL of these websites, visible comparisons of the clone and the genuine websites, crowdsourcing(based on user ratings), the use of optical character recognition, etc.
CHAPTER 3
DATASET GENERATION FOR CAMPUS SAFETY RANKING

The Clery-Act is one of the USA's most important measurement procedures for campus safety. According to the USA federal law colleges and universities are required to publish a public annual security report that provides transparency around campus crime policy, statistics, and the practices that are considered to improve campus safety. This thesis chapter focuses on the extraction of the fundamental data from the Clery-Act report, which is considered a main component in campus safety and security measurement. In our technique, we create a Python program for automating the data extraction process, which is later used to classify campuses based on their campus safety conditions and evaluate the campus safety and security rank.

3.1 Introduction

Several studies highlight the shade on campus safety and security due to the importance of maintaining a safe and productive learning environment for everyone. The Clery Act is a low feral act that guarantees that people are aware of campus crime. The Clery Act originally named crime awareness and campus security in 1990. In 1998 the act was renamed 1998 as Clery-Act. Corresponding to the Clery Act, universities and colleges receive federal funding to disseminate a public annual security report to staff, faculty, and students every year. The annual security report includes but is not limited to comprehensive information about campus statistics crime, detailed information regarding the efforts, and the higher institutions’ applied mechanism to improve campus safety and security. The annual report includes crime categories, crime reporting geography and availability, timely warning and emergency notifications, victim rights, and available Resources.

Maslow’s hierarchy describes the essential human needs. According to Maslow’s Hierarchy, the most important need for humans is a safe environment after meeting their physiological needs. Students and individuals act in fear and anxiety when the safety of the environmental conditions is not met, thus the students, faculty, and staff in the higher education environment must reach their full potential and
achieve their goals. Since the Clery-Act annual security report is a document containing the about statistics crime, campus safety procedures, resources, and campus-applied policy to reduce campus violence and increase campus safety, institutions that are participating in financial aid programs are obligated to publish the safety report information to the public to notify the individuals about campus safety.

In this research, we introduce a campus safety evaluation model utilizing the Clery-Act safety factors and measurements. This research can contribute to developing and improving safety strategies in higher institutions, by analyzing several factors that can enhance safety in higher institutions and finding the gap between high-ranking campus safety and low-campus safety. The campus safety evaluation modeling can give deep insight and understanding of the importance of safety development and guide leader and administrators to improve their safety practices by looking at and comparing sets of safety factors and measurements utilized by high-ranking institutions. Campus safety evaluation modeling can facilitate comparative studies over various institutions and point out the best practices in campus safety, analyzing safety data and factors from multiple campuses.

3.2 Background and Related Work

There is a massive amount of research about campus safety and security due to the importance of student safety and the protection of the massive institutional data collection that the university acquires. Higher institutions are considered the number one supplier for developing countries. Campus safety is significant because employees, faculty, and students feel safe, thus increasing their productivity and protecting and activating their academic and personal goals. Defining the outline for campus safety has other benefits for the community safety and the police report to study and eliminate any safety violation that can lead to unsafe and ensure campus.

Campus safety information is also important because it will empower individuals and educate them on the importance of maintaining their safety in the campus environment, helping them prepare in situations of emergency and crises off campus. It is also extremely important for the decision-making of
the campus authority to increase overall campus safety. Another reason campus safety information is important is that it is a main requirement for institutions to comply with regulations according to federal law. According to the United States Clery-Act Mandate, universities and colleges are required to report crime statistics and safety policies to comply with legal requirements. This ensures that institutions provide the necessary information for students, parents, employees, communities, and the general public.

Many people use campus safety information, including students, Parents and Guardians, faculty and Staff, campus administrators, and Security Personnel. Students depend on this information to be aware of the potential safety risk and to understand the potential campus safety risk measurements as well as any other information that helps students increase their safety while on campus. Higher institution education is a transition for college students and raises many concerns for parents choosing the right institutions for their children. These concerns can include a list of factors such as finances, location, and campus safety. Parents use this information to select the right university for their young adults. Faculty and Staff members use the campus safety information to understand the safety technique and emergency response utilized by higher institutions for their safety and spread awareness among students. Campus administrators such as university presidents, provosts, and other administrative staff utilize campus safety to develop the existing campus policy and implement novel ideas to ensure the campus community's safety.

There is no one primary resource that combines all data that can be utilized for campus safety and security research. Several reliable and trustworthy sources can be utilized for Campus safety and security, including the U.S. Department of Education, colleges and universities websites, campus police and security department, universities, and colleges ranking websites, etc. The U.S. Department of Education’s website provides access to a massive database named campus safety and security data analysis cutting Tool to provide quick and customized reports for public inquiries according to information related to fire data and campus crime.

The data analysis cutting tool allows individuals to compare different campus crime statistics across different institutions in the United States. Colleges and universities’ websites are required to
publish information about the annual security report (ASR), also known as the Clery-Act. The report includes information about the universities and colleges’ safety policies, procedures, Campus crime statistics, and other security measures and factors as well. Campus police and security department provide safety measures and crime statistics information on the university’s websites. Universities and colleges’ safety ranking websites include information about campus safety and academic quality.

Campus safety data is important to learn about the safety of campuses and assess their safety conditions and the need for further safety activities and procedures. The most available resources on campus safety provide qualitative details about various safety factors and measures considered by the educational institute as usually described on the campus website. Some agencies provide rankings published over the web and classify campuses based on their safety procedures or crime rates. However, there is no explicit ranking method that quantifies the level of compliance of campuses with the standard safety measures. A quantitative safety measure is essential to easily compare different campuses and learn about specific campus safety conditions.

In this chapter, we utilize Clery-Act reports to automatically generate important datasets about the main safety considerations of campuses based on the taxonomy described in Chapter 1 of this thesis. Our work in this chapter is the first and only automatic campus safety data generator from Clery-Act annual reports. It will help future researchers to construct new models for assessing campus safety and enhancement mechanisms. It will also automate the process of extracting and highlighting important campus safety information for concerned agencies and stakeholders. Using our work in this chapter, Campus safety can easily be evaluated every time a new Clery-Act report is published about an academic institute. Spreading and disseminating the information about the latest campus safety measures will be helpful to students, parents, faculty, teachers, and academic institute employees and their families.

3.3 Data Gathering Mechanism

In this section, we describe the data-gathering techniques to contribute to and meet a high safety and campus safety and security standard. First of all, we learned about Clery-Act’s contribution to campus
safety and security and studied the effectiveness and contribution of applying this safety measurement and factors to campus safety. Based on that, we started grouping the safety factors and measurements into groups and subgroups. We gathered over 141 Clery-Act data files in PDF form from USA universities and colleges from USA websites and transferred PDF files into text files for the data extraction mechanism.

Figure 16: Data Extraction Model Overview.

Figure 16 illustrates our technique for campus safety data extraction. In this model, First, we defined the list of campuses that will be used for this experiment and collected the Clery-Act reports for over 141 campuses of USA universities and colleges in the PDF format. Second, we created a list of the campus safety and security factors and measures based on the Clery-Act and other safety factors that universities and colleges practice in the USA to maintain a safe campus environment for the campus’s communities. Third, we collected other campus safety data including campus safety ranking. Based on the collected information, we created campus safety information which is used to generate structured campus safety data.

3.3.1 Clery-Act report definition and Structure

The Clery Act is a comprehensive report generated by institutions to meet the federal requirement of the act. The report can be compiled and presented in the following forms PDF, web-based formats, and Word documents. The Clery-Act index Introduction, statistical data and the applied campus procedure and
policy, and compliance statements are generally used. The Introduction section contains information on the main purpose of the Clery-Act, a general view or summary of the Clery-Act requirements, and the institution’s commitment to a safe campus.

The statistical data and the applied campus procedure and policy section include statistical data, charts, textual data, geographic data, time-based data, fire safety data, and data about the policy and procedures utilized by the campus authority.

The statistical data representing the statistical number of crimes that take place in the campus ground and surrounding area includes the time. Geographic data provides descriptive data, including the location and area of crime that occurred on campus. Time-Based Data information related to a reported crime, implementation of new safety facts, and the information of the date and time of reported emergency events. The policy and procedures data provide details about the utilized institutional techniques for campus policy and procedure. Fire safety data included statistical information and applied procedures for fire safety such as fire drills. According to the Clery-Act, universities, and educational institutes must include four distinct categories in their annual report: criminal offenses, hate crimes, violence against women offenses, and campus arrests and referrals for disciplinary action. Compliance statements represent the institution’s compliance with the Clery Act, the contact information of the compliance officer, and information regarding campus safety and resources.

Table 3- Sample Clery-Act report contents - Western Michigan University Clery-Act report 2021 is a sample of the Clery-Act report contents from Western Michigan University. From the sample image, we can see a list of the campus safety factors and measures which is utilized across the higher institutions of the United States, for example, crime prevention, alcohol, and drug policy, missing students’ notification policy, crime statistics by campus, sexual assault policy, emergency notification, etc.
Higher institutions that participate in federal funding are required to publish a Clery-Act report to students and employees every October first. The report contains statistical info about the university and college campus crime for the previous three years. The safety and security report also contains different information related to the implemented policy, the emergency procedure, and others.
3.3.2 Campus Safety Ranking Data.

Table 4- Sample of campus safety ranking for a list of universities and colleges across the United States from College Transitions [20].

<table>
<thead>
<tr>
<th>Institution</th>
<th>On-Campus Safety Index</th>
<th>Off-Campus Safety Index</th>
<th>Overall Safety Index</th>
<th>Off-Campus Safety Rank</th>
<th>On-Campus Safety Rank</th>
<th>Overall Safety Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bluefield State College</td>
<td>5</td>
<td>3.5</td>
<td>5</td>
<td>316</td>
<td>1</td>
<td>48</td>
</tr>
<tr>
<td>Cleveland Institute of Art</td>
<td>5</td>
<td>3.5</td>
<td>4.5</td>
<td>726</td>
<td>1</td>
<td>129</td>
</tr>
<tr>
<td>Granite State College</td>
<td>5</td>
<td>3.5</td>
<td>3</td>
<td>343</td>
<td>1</td>
<td>57</td>
</tr>
<tr>
<td>Grove City College</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>586</td>
<td>1</td>
<td>42</td>
</tr>
<tr>
<td>Hillsdale College</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>379</td>
<td>1</td>
<td>38</td>
</tr>
<tr>
<td>New York Institute of Technology</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>18</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>St. Josephs College-Long Island</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>St. Josephs College-New York</td>
<td>5</td>
<td>2</td>
<td>4.5</td>
<td>640</td>
<td>1</td>
<td>176</td>
</tr>
<tr>
<td>Stevenson University</td>
<td>5</td>
<td>4.5</td>
<td>5</td>
<td>96</td>
<td>1</td>
<td>8</td>
</tr>
<tr>
<td>SUNY Empire State College</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>55</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>United States Air Force Academy</td>
<td>5</td>
<td>0.5</td>
<td>4.5</td>
<td>929</td>
<td>1</td>
<td>100</td>
</tr>
<tr>
<td>United States Coast Guard Academy</td>
<td>5</td>
<td>1.5</td>
<td>5</td>
<td>738</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>United States Military Academy</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>50</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>United States Naval Academy</td>
<td>5</td>
<td>0.5</td>
<td>5</td>
<td>862</td>
<td>1</td>
<td>82</td>
</tr>
<tr>
<td>University of Maryland-University College</td>
<td>5</td>
<td>2</td>
<td>4.5</td>
<td>567</td>
<td>15</td>
<td>147</td>
</tr>
<tr>
<td>CUNY Brooklyn College</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>223</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>Nevada State College</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>213</td>
<td>17</td>
<td>22</td>
</tr>
<tr>
<td>CUNY Bernard M Baruch College</td>
<td>5</td>
<td>1</td>
<td>1.5</td>
<td>813</td>
<td>18</td>
<td>670</td>
</tr>
<tr>
<td>Thomas Aquinas College</td>
<td>5</td>
<td>3.5</td>
<td>5</td>
<td>349</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td>College of the Ozarks</td>
<td>5</td>
<td>2.5</td>
<td>5</td>
<td>486</td>
<td>20</td>
<td>34</td>
</tr>
<tr>
<td>CUNY Lehman College</td>
<td>5</td>
<td>2</td>
<td>3.5</td>
<td>579</td>
<td>21</td>
<td>370</td>
</tr>
<tr>
<td>The University of Texas Rio Grande Valley</td>
<td>5</td>
<td>2.5</td>
<td>4</td>
<td>494</td>
<td>22</td>
<td>265</td>
</tr>
<tr>
<td>Kettering University</td>
<td>5</td>
<td>1</td>
<td>2.5</td>
<td>818</td>
<td>23</td>
<td>484</td>
</tr>
<tr>
<td>Mississippi University for Women</td>
<td>5</td>
<td>1.5</td>
<td>4</td>
<td>683</td>
<td>24</td>
<td>235</td>
</tr>
<tr>
<td>University of Michigan-Dearborn</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>226</td>
<td>25</td>
<td>30</td>
</tr>
<tr>
<td>College for Creative Studies</td>
<td>5</td>
<td>1</td>
<td>3.5</td>
<td>753</td>
<td>26</td>
<td>288</td>
</tr>
<tr>
<td>Metropolitan State University</td>
<td>5</td>
<td>3</td>
<td>4.5</td>
<td>430</td>
<td>27</td>
<td>94</td>
</tr>
<tr>
<td>CUNY Hunter College</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>790</td>
<td>28</td>
<td>636</td>
</tr>
<tr>
<td>CUNY John Jay College of Criminal Justice</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>796</td>
<td>29</td>
<td>647</td>
</tr>
<tr>
<td>University of Baltimore</td>
<td>5</td>
<td>0.5</td>
<td>3</td>
<td>864</td>
<td>30</td>
<td>383</td>
</tr>
<tr>
<td>Utah Valley University</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>196</td>
<td>31</td>
<td>21</td>
</tr>
</tbody>
</table>

College Transitions is a website that includes information about campus safety ranking from a list of universities and colleges across the United States. The data is in the form of a comprehensive array of tables and lists created to help students and their families navigate colleges and compare different information that guides students to prepare for a final academic decision. The website includes information about campus safety ranking for a list of colleges and universities in the U.S. The campus data is collected from the Department of Education. The campus safety ranking includes different ranking categories such as campus safety index, off-campus safety index, overall safety index, off-campus safety...
rank, and overall safety rank. The star rating system is used to compare institutions to their peers. Table 4 is a sample of campus safety ranking data from the collected sample of 141 campuses for a research experiment aiming to assess and compare the safety levels among higher institutions, and this comparison is based on numerous safety and security factors and crime rates.

3.4 Data Extraction Model Overview

Figure 17 - Data extraction from Clery-Act reports and dataset generation model.

Figure 17 Data extraction from Clery-Act reports and dataset generation model explain the techniques utilized for data extraction utilizing the created Python code for the data extraction for the text file. The Clery-Act for universities and colleges is converted to text file format in the first step. The next step is the campus safety report fetch of campus safety keywords and the similarity of groups of safety
terms. Finally, form the created text file utilizing the Python code for the data set extraction generation. If the safety term and others similarly exist in the text file that will indicate the number of safety occurrences of the classified safety term in the Clery-Act text file.

3.5 Input Data Preparation

We Convert PDF files into Text using `pdftotext` from XpdfReader. XpdfReader is a free PDF file reader tool for operating systems based on the Qt toolkit [https://www.xpdfreader.com/download.html]. The reader provides a set of useful tools for processing, reading, and converting pdf files, among several other tools. Of these tools, we use the pdf text converter (`pdftotext`) which converts pdf files to text files by removing binary contents such as images, colors, and text formatting. The `pdftotext` tool can be used in the command prompt, and thus it can be programmed as part of shell script code that executes the commands on a number of input pdf files and converts all of them using a system operation. The following figure (Figure 18) describes a simple call of the `pdftotext` on the command prompt for converting a file called `CleryActReprt.pdf` and saving the result text file as `CleryActReprt.txt`.

```
pdftotext CleryActReprt.pdf CleryActReprt.txt
```

*Figure 18- Pdftotxt command sample.*

Figure 19 shows a shell script for converting all pdf files on a specific folder (e.g., D:\GS\Thesis\CampusReports\PDFs\) into text files and saving them in the same folder. As shown in the figure, we use for loop command by defining an iterative variable (%F) that represents a pdf file from the specified folder at each iteration. The variable %F is used as a parameter for the pdftotext command on the DO section of the FOR command. With each iteration, the pdftotext will convert a pdf file from the folder into a text file with the same name.

```
FOR %F IN ("D:\GS\Thesis\CampusReports\PDFs\*.pdf") DO pdftotext "%F"
```

*Figure 19- Sample shell script for converting multiple pdfs into text files.*
Figure 20 shows the shell script command on the command prompt (shell program). It also shows a snapshot of the results of executing this command line by line.

3.6 Specifying Extracted Data Structure

Our model specifies a set of key terms based on our taxonomy described in Chapter 1 of this thesis. These key terms consolidate the major campus safety factors and measures as described in the Clery-Act reports of many (exactly 141) high educational institutes ranked from lowest to highest on the major ranking websites. The key terms are compiled on a text file with a specific structure that enables our data extraction model to search these key terms in Clery-act annual reports. The key terms are grouped into main campus safety evaluation components that categorize each group of factors of the same scope or concern. For each key term, we specify a number of aliases. Aliases are other words or expressions that have the same or close meaning with a minor or major difference in words, spelling, or term structure. For example, the key term: emergency drills have the following aliases:

- Drills
- Evacuation procedure
- Emergency Evacuation
- Emergency Training

The existence of any of these terms in the Clery-Act report would count the same way towards key-term Emergency drills. Aliases enable the consideration of different writing styles of the different
Clery-Act reports created by different writers. For example, the existence of the “Camera surveillance” term or any of its aliases will indicate that the campus Clery-act report addresses the security Camera system in the campus spaces.

Figure 21 - Safety term groups and aliases.

Figure 21 shows a sample of the key terms file that we use in our dataset generator. The sample key term file describes nine main safety categories, including annual security reporting, emergency management, security policy, … etc. Under each of these categories, we specify a number of keywords or key terms that represent a specific standard campus safety procedure, activity, or measure. We generate
this file manually and tune it to the best set of safety categories, safety measures, and their aliases of campuses. We can define any number of aliases as needed for each key term. Using the colon (“:”) character, we separate each key term and its aliases. We specify a campus safety category by preceding it using the at character (“@”). In this text file, a new line (Enter or carriage return) introduces a new campus safety category or a campus safety key term. This file ignores empty lines while parsing the key terms in our dataset generator.

3.7 Data Extraction Model Implementation

Our data extraction model combines a set of input reports and data files to generate the desired data set specified by the key terms and groups listed in the key terms text file. The inputs for our extraction model are the list of annual Clery-Act reports generated by the campuses, the generated key terms file, and the latest ranking based on any trusted publishing agency. Here, we use the college transition website due to the availability of numerous campus safety on it, and the availability to download their dataset with multiple ranking methods applied for each campus. They also consistently update their information based on the educational institutions and their published information [20].

![Campus safety data extraction and dataset generation model overview](image)

As shown in Figure 22, our data generator is created in Python language. We use Python 3.9 with Microsoft visual studio code 1.78. Our program is made of 323 lines of code distributed over 10 program
functions. In the figure below (Figure 23), we show part of the code, the main function that illustrates the main flow of our program.

```python
    def main():
        term = []
        group = []

        postfix = outDs
        skip = 1
        for c in CampusEvaluationFile:
            if skip == 1:
                skip = 0
                continue
            L = c.split(',
            campus.append(L[0])
            campusEvaluation.append(L[:])

        CampusEvaluationFile.close()

        save = open('Dataset' + postfix + '.csv', "w")
        print("--- Reading terms and term groups ........")
        term, group = getTermsAndGroups()
        writefileHeader(save, term, group)

        print("--- Creating Dataset file ........")
        term, group = termsToLower(term, group)

        Linel = []
        RankL = []
        lineL, RankL = createDataset(RankL, Linel, term, group)

        writeDataset(save, linel)

        print("--- Dataset is generated ........")
        save.close()

    if __name__ == "__main__":
        main()
```

Figure 23- Main functions of Python implementation of Campus Safety dataset from Clery-Act reports.

3.8 Sample Campus Safety Dataset Results

The output of our campus safety data generator is mainly comma-separated text files, or more than one file based on a specific parameter for specifying the output file formats, headers, contained campus safety measures, and statistical analysis. Our data generator program is also capable of creating separate data files for different settings of the above-mentioned configurations. Some of these configurations are related to the campus safety data classifier and will be explained in the next chapter.

The following figures (Figure 24 and Figure 25) show samples comma separated values (CSV) files.

Figure 24 shows a sample dataset based on the safety term occurrence counter in campus Clery-Act
Reports, while Figure 25 shows a sample dataset based on the Safety term existence indicator in Campus Clery-Act Reports. In these two output versions, the occurrence counter indicates how many times a term and its aliases have occurred in the Clery-act report of a specific campus. The existence indicator is a Boolean value (True or False) or (1 or 0) indicating whether a term or its aliases exist in the report at least a few times.

**Figure 24- Sample dataset based on Safety term occurrence counter in Campus Clery-Act Reports.**

Other out data format that our program generates is WEKA-ready, i.e., a structured data model for the WEKA data mining tool. In this format, we prepare our WEKA-ready (i.e., arff file) data file to be run automatically on Weka without any need to specify the data formats or headers in WEKA itself. We also prepare our data by normalizing all fields to the range of 0 to 1. To facilitate the regression process with minimal effort on the WEKA tool. This allows us to run WEKA prediction analysis for several regression
models on datasets generated by our extraction program without effort. As shown in Figure 26 the generated data file is structured and ready for WEKA by adding the header, and field attributes, data types, and system keywords. Figure 27 shows our WEKA-ready file opened in WEKA by only double clicking, where the system is ready to start data analysis procedures without needing any data preparation.

Figure 25- Sample dataset based on Safety term existence counter in Campus Clery-Act Reports.
Figure 26- Sample dataset prepared automatically and structured for Weka data mining tools.

Figure 27- A sample dataset automatically created for Weka when opened on the Weka editor by double-clicking the file.
3.9 Summary of Dataset Generation for Campus Safety Ranking

In this thesis chapter, we provide an extraction model for campus safety data from their annual Clery-Act reports. The extracted dataset is important for campus safety and security measurement, classification, and ranking. In our data extraction technique, we create a Python program for automating the data extraction process, which is later used to classify campuses based on their campus safety conditions and evaluate the campus safety and security rank.

Our technique creates a number of data formats and specifies the data contents based on system configurations. Examples of output dataset formats include comma-separated values (CSV) files, and arff (WEKA-ready) files. Our program can also create a list of separate files based on the provided system configurations, where each file represents a specific campus safety class. Our program can also generate datasets based on the safety term occurrence counter or Safety term existence indicator in Campus Clery-Act Reports.
CHAPTER 4

AUTOMATED CAMPUS SAFETY CLASSIFICATION AND RANKING

Campus safety classification is important for contrasting different campuses based on their safety measures. On the College Transitions website, Campuses are ranked based on six different measures, as described in Error! Reference source not found. Some of their campus safety rankings depend on an index factor that quantifies campus safety as an index from 1 to 5-star rating. Such types of ranking provide a classification of campuses based on the degree or level of campus safety measures from low to high. According to College Transitions, the 0.5 to 5-star rating system is generated by comparing institutions to their peers. For example, “those with a 5-star safety index fall within the top 10% of institutions in the dataset. Additionally, an "Overall Safety Index" was generated, including on-campus and off-campus safety measures. How heavily the on-campus and off-campus crimes were weighted depends on how residential the institution is, using the IPEDS dorm capacity variable. A highly residential institution (e.g., Ursinus College) has an on-campus safety index of only 0.5 stars (i.e., less safe) and an off-campus crime index of 4.5 stars (i.e., very safe). Their overall safety index is 1 star since most students reside on campus. Meanwhile, despite Youngstown State (a highly commuter university) having a 4.5-star on-campus safety index, their overall safety index is only 2-stars given the high crime rates in Mahoning County, Ohio” [20].

In our generated dataset, we noticed that some campus datasets are not producing a high correlation compared to the provided ranking on College Transitions; others produce a high correlation coefficient. Considering the facts about the campus safety index on College Transitions, we realize that the campus neighborhood has a major influence on the campus safety index. While campus safety procedures are major contributors to the overall safety of the campus, campus safety procedures have partial or minor influence over the campus safety index on College Transitions. To provide high emphasis on campus safety procedures on their campus safety rank, we attempt to classify campuses into three main classes based on the variance between safety procedures and safety influence from the campus
neighborhood. We call these classes: positive, neutral, and negative neighborhood influence on campus safety classes. For short we refer to these classes as positive, neutral, and negative classes and we annotate them as class -1, class 0, and class 1: for positive, neutral, and negative classes, respectively. Therefore,

- **Class 1**: positive neighborhood influence over the campus ranking. These campuses are usually ranked high safety and security based on their safe location, disregarding the low safety procedures considered on these campuses.

- **Class 0**: neutral neighborhood influence over the campus ranking. These campuses are usually ranked mainly by considering the campus safety conditions and procedures to secure daily activities and avoid harmful threats on campuses.

- **Class -1**: negative neighborhood influence over the campus ranking. These campuses are usually ranked low in safety and security based on the unsafe or hostile location disregarding the high campus safety procedures considered on these campuses.

Our work in this chapter also helps predict the safety rank of campuses based on their Clery-Act report by comparing it to existing Clery-Act reports of other campuses as well as the reported rank on public resources. The following sections explain the detailed description and results of our classification and prediction models.

4.1 Model Overview

Our model reads as input the Clery-act-based campus safety dataset extracted in the previous Chapter. Based on the Campus rank in College Transitions and its safety data measures and factors extracted based on the key term file, our classifier is written in Python 3.9, and it classifies each campus into one of three classes specified earlier in this chapter. We classify existing campuses based on their safety conditions and conformance with the level of crime rates as ranked on public websites (here, we use College Transitions ranking).
4.2 Classification Model

This section describes our classification model based on the safety condition conformance with the crime rate levels published on public websites. The detailed process of our classification method involves calculating several measures for each campus based on its reported safety measures and crime rate ranking. These measures are defined in the following Table.

Table 5: Statistical data measures used in our classification process.

<table>
<thead>
<tr>
<th>Classification process measure</th>
<th>Definition and explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DCSM</strong>: Density of Considered Safety Measures</td>
<td>The DCSM is the number of safety measures and activities a specific campus conducts relative to the total number of measures considered in the dataset generation. For example, 44 measures can be considered in the data extraction for each campus. Thus, the best campus would consider the one that considers all 44 measures and the lowest is the one that considers zero measures of these. Any number of safety measures between 0 to 44 considered by a specific campus is called the Density of Considered Safety Measures (DCSM)</td>
</tr>
<tr>
<td><strong>RND</strong>: Rank-Based Normal DCSM</td>
<td>The Rank-Based Normal DCSM (RND) is the conforming DCSM number based on the campus safety rank in the public agencies. For example, if a campus rank is 2.5 stars, then should consider about half of campus safety measures assuming all measures have the same weight on the campus safety risk evaluation. Here, if 44 measures are considered then a campus of 2.5 stars is expected to ideally have RND of 22 measures considered. If a campus DCSM = RND. That means this campus is typically evaluated based on the safety measures and procedures and it almost has no influence from the campus neighborhood. If a campus DCSM &gt;</td>
</tr>
</tbody>
</table>
RND, that means this campus has some negative influence from the campus neighborhood. Similarly, if a campus DCSM < RND, that means this campus has some positive influence from the campus neighborhood.

**ADM:** Accepted Density Margin
A configuration threshold of the accepted deviation from the Rank-based normal DCSM.

**NDR:** Normal Density Range
The Normal Density Range (NDR) is the interval of all the values from RND-ADM to RND+ADM.

**Min-ND:** Min Normal Density
Min-ND is the value RND-ADM

**Max-ND:** Max Normal Density
Max-ND is the value RND+ADM

**DCSM Deviation**
DCSM Deviation is the distance between a campus DCSM and the NDR range. For example, if the campus DCSM is 23 and the NDR range is [8-12] then the DCSM deviation is 23-(Max-ND). I.e., DCSM deviation = 23 – 12= 11.

If the campus DCSM is 23 and the NDR range is [24-32] then the DCSM deviation is 23-(Min-ND). I.e., DCSM deviation = 23 – 24= -1. The DCSM deviation is the final measure that specifies the campus class. Positive DCSM deviation means the campus class is 1, negative DCSM deviation means the campus class is -1, otherwise, 0 DCSM deviation means the campus is in class 0.

Table 5- Statistical data measures used in our classification process. Table 5 illustrate the meaning of each measure in our campus safety classification process. The following figure shows how these measures are captured and aggregated towards a decision about campus safety class as described early in this chapter.
Figure 29—Campus safety classification model details.

Figure 29—Campus safety classification model details shows the flow of calculations to aggregate a campus safety class from the input dataset. We first calculate the DCSM based on the number of safety measures considered in a specific campus. We then read the campus safety rank calculated in the public agencies, and then we use to calculate the RND, NDR, Min-ND, and Max-ND by using the ADM setting (say 4). Our experiment shows that ADM = 3 provides the best classification results, as explained later in this chapter. We then compare DCSM to the NDR (by comparing to Min-ND and Max-ND) and calculate the DCSM deviation. Finally, we specify the campus safety class.

Figure 30—Three campus safety classes based on campus safety data and current ranking index.

Figure 30 describes a high-level overview of the classification model, particularly the last step of deciding the campus safety class based on the calculated DCSM and DCSM deviation.

4.1 Campus Safety Classification Results

Table 6 provides a complete example of campus safety classification by showing all calculations and aggregations of classification process measures for several campuses (here, 60 campuses). The table
shows a group of campuses in each campus safety class 1, 0, and -1. For example, Augusta University belongs to category (class -1) by a DCSM deviation of -1 safety measure. I.e., the campus has one less safety measure than its normal DCSM according to the public safety rank. Kennesaw state university considers 40 of the addressed 44 safety measures and since it is ranked 4.5 stars on College Transitions, it is already classified in our model as class 0. This also indicates that Kennesaw state university has minimal influence from the campus neighborhood on its public safety ranking or that it is taking exact safety measures as conforming with its public rank despite its neighborhood safety.

Table 6- Classification model calculation sample data from different safety classes.
As an example of a class 1 campus, we use Emory University, which is ranked 1.5 on College Transitions, while it considers 36 campus safety measures. But due to its low rank on College Transitions, it is classified as class 1: a campus that takes more safety measures than its peers of the same safety rank on the public ranking. This is due to the neighborhood’s negative influence on its rank.

In the above table, we use Accepted Density Margin (ADM) = 3, i.e., the NDR is a set of 7 acceptable DCSM values. If a campus DCSM belongs to this NDR range, then it is classified as Class 0. Thus clearly, the ADM settings control the number of campuses classified in class 0, and consequently, it affects the number of campuses in the other classes as well. A larger ADM means a bigger NDR and more campuses in class 0, and consequently, fewer campuses in class -1 and class 1. The following Figure illustrates the relationships between ADM and the number of campuses in each different class.

![Figure 31-Number of Campuses per safety class by Accepted Density Margin.](image)

Figure 31 shows the ADM setting on the x-axis and the number of campuses on the y-axis. It is clear that, with the increase of ADM, class 0 increases too. Since class 0 increases in the number of campuses, the other two classes will decrease as shown in the bar chart.
4.2 Campus Safety Ranking Model

Here we introduce our campus safety ranking model that can predict safety rank for campuses that do not have a publicly available ranking. We use the campus safety measures and public ranking of peer campuses with similar safety conditions and procedures. Thus, our model aims to train a model to predict campus safety based on a specific dataset. We use our campus safety dataset created in the previous chapter. We use several regression analysis methods implemented in the WEKA data mining tool to train our campus safety ranking model as described in Figure 32. The following figure illustrates the details of our campus safety ranking model.

Figure 32- Campus safety prediction model overview.

Figure 33- Campus safety prediction model implementation.

Figure 33 illustrates the details of our campus safety ranking. Our prediction model utilizes the generated dataset of campus safety measures as described in the annual Clery-Act report of different campuses. We then use a Python program to generate a WEKA-ready data file, as explained in the
previous chapter. The WEKA-ready files then run-on WEKA, and we choose the regression model for our already normalized dataset. The independent variables are campus safety measures (44 fields) as specified by our key term file. We run several regression models against various campus safety classes (Class -1, class 0, and Class 1, as specified earlier). We also generate our dataset with multiple data files for a range of ADM setting values. Figure 34 shows some of the input data files imported into WEKA for regression analysis using five main models as follows.

- Gaussian process
- Linear regression
- Multilayer Perceptron
- Simple Linear regression
- Random Forest

![Figure 34 - Sample WEKA-ready (.arff) data files for different classes and different ADM values.](image)

We compare the regression analysis methods with our model based on the correlation coefficient and other statistical measures, as shown in the results section later in this chapter. The regression analysis split the input data into two groups: the training group, 66.6% of the input data, and the simulation (or testing) group, 33.3% of the input data. The testing data is not used for training; thus, they are considered new samples for the regression system. Predicting results of these new samples represent the correlation coefficient of our regression analysis model.
4.3 Campus Safety Ranking Results

This section presents the results of our campus safety ranking model. The following table shows the main statistical measures of several regression models by utilizing different data files for each campus safety class. Here we use ADM = 3 for the acceptable density margin.

<table>
<thead>
<tr>
<th>Model</th>
<th>all data</th>
<th>Class (-1)</th>
<th>Class (0)</th>
<th>Class (1)</th>
<th>all data</th>
<th>Class (-1)</th>
<th>Class (0)</th>
<th>Class (1)</th>
<th>Total Number of Instances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gaussian Process</td>
<td>0.1549</td>
<td>0.3475</td>
<td>0.9047</td>
<td>0.3865</td>
<td>0.1711</td>
<td>0.7985</td>
<td>0.3038</td>
<td>0.1017</td>
<td>48</td>
</tr>
<tr>
<td>Linear Regression</td>
<td>0.254</td>
<td>0.0813</td>
<td>0.0479</td>
<td>0.1303</td>
<td>0.2052</td>
<td>0.0622</td>
<td>0.1494</td>
<td>0.0875</td>
<td>13</td>
</tr>
<tr>
<td>Multilayer Perceptron</td>
<td>0.3019</td>
<td>0.1073</td>
<td>0.0575</td>
<td>0.1613</td>
<td>0.2914</td>
<td>0.074</td>
<td>0.1818</td>
<td>0.1017</td>
<td>10</td>
</tr>
<tr>
<td>Simple Linear Regression</td>
<td>0.108057</td>
<td>0.929247</td>
<td>0.507331</td>
<td>0.910697</td>
<td>1.043298</td>
<td>0.658485</td>
<td>1.044346</td>
<td>0.647679</td>
<td>26</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.0288</td>
<td>1.09288</td>
<td>0.5032</td>
<td>0.970177</td>
<td>1.054781</td>
<td>0.647679</td>
<td>1.093944</td>
<td>0.647679</td>
<td>26</td>
</tr>
</tbody>
</table>

It is clear in Table 7- Prediction precision by regression model for different campus safety classes that, for all regression models, class 0 has the highest correlation coefficient among all other classes, it is also always higher than the correlation coefficient of all data together in one WEKA file. For example, linear regression provides a correlation coefficient of 0.7985 for class 0 data where the total number of instances (testing samples) is 10. For linear regression, the other classes provide 0.1711, 0, and 0.3038 for the class of all data, the class (-1), and the class 1, respectively. The same pattern of numbers
repeats with the other regression models. Similarly, the correlation error measures such as the mean absolute error, the root means square errors, and the other error measures in the table are notably lower for class 0 and higher for every other class of data. The data in Table 7 are plotted using bars in the following figure.

Figure 35- Prediction precision by regression model for all campus data.

Figure 35 presents the analysis results of five regression models, as shown in the figure. The regression analysis is applied for all campus data (i.e., no specific class) with an ADM setting of 3. The data of this chart is simply the first row in each regression model in the previous table.
Figure 36 - Prediction precision by regression model for class 0 with ADM = 3.

Figure 36 compares each regression model's correlation coefficient and statistical error measures with the class 0 dataset in a line chart. The Gaussian process, the multilayer perceptron, and the random forest have the highest correlation, and simple linear regression has the lowest. Error measures can also be compared in the same way.

Figure 37 - Prediction precision by regression model for class (-1) with ADM = 3.

Figure 37 compares the same values of Figure 36 by using a different dataset for class (-1). Here, we notice that the correlation coefficient is low relative to the standard error measures for all regression models. Similarly, Figure 38 provides the same data for class 1.
Figure 38-Prediction precision by regression model for class 1 with ADM = 3.

Figure 39-Corelation Coefficient by prediction model by class with ADM = 3.

Figure 39 illustrates the predictability feature of class 0 relative to the other classes. The figure plots the correlation coefficient for each regression model for each campus safety class data set (class (-1), class 0, class 1). It is also clear that all data that include all classes at the same time has the lowest correlation of all. That also proves that our classification improves the predictabilities of all classes, particularly class 0.
Figure 40 - Total Number of campuses and number of Instances used for regression for each ADM.

Figure 40 shows the number of campuses in class zero and the number of campuses used as instances for testing the regression model each time WEKA runs the dataset. These numbers vary based on the chosen ADM value represented on the x-axis. The following figures display the regression analysis results for each regression model in a separate chart. Gaussian process regression results are displayed in Figure 41, linear regression is presented in Figure 42, Multilayer perceptron is presented in Figure 43, simple linear regression is presented in Figure 44, and finally, random forest regression analysis data is represented in Figure 45.

Figure 41-Gaussian Process Regression by ADM for Class 0.
Figure 42 - Linear regression by ADM for Class 0.

Figure 43 - Multilayer Perceptron regression by ADM for Class 0.
4.4 Summary of Campus Safety Classification and Ranking

In this Chapter, we utilize our campus safety data generator described in Chapter 3 to classify existing campuses based on their safety conditions. Our classification is based on the neighborhood influence on the public campus rank despite the campus safety conditions and activities contributing to safer learning and working environments. The classification based on neighborhood influence on campus safety rank helps distinguish classes of campuses with positive, neutral, and negative neighborhood influence on campus safety. For short, we refer to these classes as positive, neutral, and negative classes.
and we annotate them as class -1, class 0, and class 1: for positive, neutral, and negative classes, respectively.

The above classification enables our campus ranking or rank prediction model to control the level of influence in the campus safety ranking mechanism by simply choosing the campus safety class for the regression analysis model. To provide high emphasis on campus safety procedures with no neighborhood influence on their campus safety rank, we use the class 0 dataset to train our ranking model. Our campus safety ranking model helps predict the safety rank of campuses based on their Clery-Act report by comparing it to existing Clery-Act reports of other campuses as well as the reported rank on public resources.

CHAPTER 5
CONCLUSIONS

Campus safety measurement and ranking are important information for many stakeholders, including students, families, faculty, teachers, and educational domain workers. However, existing campus safety rankings are mostly based on crime history with no consideration of campus safety and security conditions and activities. The Clery-Act report is considered the US's most important campus safety report. Most US universities and colleges prepare an annual report for their safety standards and crime incidents during the reporting interval. In this thesis, we utilize Clery-Act reports of campuses to automatically analyze their safety conditions and generate a safety rank based on these reports. We cover a specific research activity in each chapter of this thesis as follows.

Chapter 1 provides a survey of main campus safety measurement components and factors that are utilized in the annual Clery-Act reports of academic institutes. We explain each safety factor and discuss its details in this chapter. We also describe the distributions of these factors in the annual Clery-Act reports among some other results.

Chapter 2 highlights the main research contribution to campus safety enhancement and assessment activities that attempts to improve campus safety conditions with a specific safety concern and/or specific student’s (or human) environment. This chapter also describes specific safety threats on
campuses, specifically in academic institutes. We describe the campus security landscape and illustrate the level of threat on these campuses, and then we describe the protection mechanisms.

In Chapter 3, we construct a mechanism for campus safety dataset generation based on the annual Clery-Act report of campuses. Our data-gathering mechanism utilizes the taxonomy provided in the Chapter to define a set of key terms and safety categories/groups. These terms are used to search the annual Clery-Act reports for related information. Our campus safety data generator can generate comma-separated values (CSV) files or more than one CSV file based on a specific parameter for specifying the output file formats, headers, contained campus safety measures, and statistical analysis. Our data generator can also produce WEKA-ready input files (i.e., a structured data model for WEKA data mining tool: .arff file). The generated data is normalized and completely ready for WEKA analysis and data processing.

Chapter 4 utilizes the collected information in Chapter 3 to classify existing campuses based on their safety conditions. We classify campus safety data into three classes (class 1, class 0, and class -1). Class 1 includes campuses that have positive neighborhood influence over their public ranking. Class 0 includes campuses that have neutral or no neighborhood influence over the campus ranking. Finally, Class -1 includes campuses that have negative neighborhood influence over their ranking. These classes also have different behavior when utilized in the regression analysis to predict the safety rank of campuses that do not have a public record for their ranking. In this chapter, we also predict the safety rank of campuses based on their Clery-Act report by comparing it to existing Clery-Act reports of other campuses and reported rank on public resources. Our results at the end of this chapter show the effective use of our classification and prediction model, specifically when used with Class 0 data with an Acceptable density margin of 3.

Our future work will focus on the following directions.

1) Campus safety evaluation based on weighted campus safety factors. In the previous research, we assumed that all campus safety factors and measures have the same impact on safety evaluation. Practically, these measures have different impacts on campus safety, and therefore, these
differences must be considered. In the future, we plan to consider these differences by studying their impact and allowing specifying the different weights of each measure on campus safety evaluation.

2) Controlled neighborhood safety influence on Campus safety evaluation. In the previous research, we conducted most of our experimental evaluation by using Class 0, where no influence from the campus neighborhood on its safety evaluation. In future research, we plan to consider partition Class 0 and the other two classes (Class 1 and Class -1) and build a training data set from these partitions based on the level of influence of neighborhood on campus safety measures.

3) Emphasized Campus security evaluation by focusing on IT issues and cyber security threats. In the previous research, we conducted all safety impacts including some managerial, security, safety, health, food, and residence safety measures. In the future, we will attempt to focus on information technology and cyber security concerns to calculate a campus technology safety measure that illustrates how safe to use the campus technology, communication, and equipment.
REFERENCES

[6]. “Public Safety at the University of Saint Joseph | USJ West Hartford, CT,” University of Saint Joseph. https://www.usj.edu/student-life/campus-safety


[47]. Writers, S. (2022b, August 26). College Student Resources and Tools. Affordable Colleges Online.


APPENDIX A

Python code for dataset generator

```python
import os

PdfTxtPath = ".\PdfTxt\"
pdfPath = os.path.dirname(os.path.abspath(__file__)) + PdfTxtPath
RepFlList = os.listdir(pdfPath)

outs = "0" # for Xs Put X, for numbers put #, for 0s, 1s put 0
outClasses = [-2,-1,0,1] # for class 0, -1, 1 Put 0, -1, 1 to the list, any number (e.g., -2) means all data
WekaReady = False # True # False # True # False#
tryADMs = [3] # for ADM 4 Put [4], else put [.....] << list of ADMs, [0,1,2,3,4,5,6,7,8]

fileClass = 0

if WekaReady:
    OutputExt = 'arff'
else:
    OutputExt = 'csv'

CampusEvaluationFile = open('CampusEvaluations.csv')
campus = []
campusEvaluation = []

def sort(LineL, RankL):
    newL = []
```
for i in range (len(LineL)):
    newL.append(str(round(float(RankL[i]),1)).zfill(3) + "-" + LineL[i])

newL.sort()
newL2 = [x[4:] for x in newL]

return newL2

def getInstEval(Inst):
    loc = campus.index(Inst)
    if loc < 0:
        print (Inst, 'not found')
        return (-1)
    else:
        if (Inst == campusEvaluation[loc][0]):
            return (campusEvaluation[loc][1]) # index 1 = On-campus safety Index (stars 0.5 to 5.0), index 5 = On-campus safety Rank (1 to 999),
        else:
            return (-2)

def getTermsAndGroups():
    theGroup = ""
    term = []
    group = []
    termsFile = open('SafetyTerms.txt')
    for t in termsFile:
        t= t.lower()
        isGroup = t.startswith('@')
        t = t.strip()
        if len(t)>0:
            if isGroup:
                theGroup = t[1:]
                continue

            if (t.endswith(':')): t = t[:-1]
            Alternatives = t.split(':')

            term.append(Alternatives)
            group.append(theGroup)
    termsFile.close()
    return (term, group)

def writeFileHeader(save, term, group):
    lineGroup = "GRP>>

line = "Report 

prvGroup = ""
for i in range(len(term)):
    if (prvGroup != group[i]):
        lineGroup += ', ' + group[i]
        prvGroup = group[i]
    else:
        lineGroup += ',

    line += ', ' + term[i][0]

lineGroup += ' ' + 'Campus Safety Ranking'
line += ' ' + 'On-Campus Safety Rank'

save.write(lineGroup + "\n")
save.write(line + "\n\n\n\n")

def termsToLower(term, group):
    for i in range(len(term)):
        for j in range(len(term[i])):
            term[i][j] = term[i][j].lower()

        group[i] = group[i].lower()

    return (term, group)

def getReportPart(Report, wFrom, wTo, backward = False):
    wFrom = wFrom.lower()
    wTo = wTo.lower()

    wToLoc = len(Report) - 1
    if (not backward):
        wFromLoc = Report.find(wFrom)
    else:
        wFromLoc = Report.rfind(wFrom)

    if (wFromLoc < 0): return ""

    if (wTo > ""):
        wToLoc = Report.find(wTo, wFromLoc, len(Report))

    if (wFromLoc>=0 and wToLoc>=0):
        return Report[wFromLoc:wToLoc]
    else:
def createDataset(RankL, LineL, term, group):
    termOccurrence = []

    for f in RepFlList:
        if f.endswith(".txt"):
            Report = str(f)[:-4]
            line = (str(f)[:-4]).replace(";", ",").replace("_", ",")

            InstEval = getInstEval(Report)

            f = open(PdfTxtPath + f)
            Report = f.read().lower()
            termOccurrence = ["" for _ in range(len(term))]
            f.close()

            for i in range(len(term)):  # All Terms
                SumOccurrence = 0
                SumMainPartsOccurrence = 0

                for j in range(len(term[i])):  # All alternatives
                    SumOccurrence += Report.count(term[i][j])
                termOccurrence[i] = SumOccurrence

                v = ""
                if (SumOccurrence > 0):
                    if (outDs=="X"):
                        if (SumOccurrence>3):
                            v = 'X'
                        else:
                            v = 'x'
                    elif (outDs=="#"):
                        v = str(SumOccurrence)
                    else:
                        v = "1"
                else:
                    if (outDs=="#"):
                        v = "0"

                if (SumOccurrence == 0 and outDs == "0"):
                    v = "0"
line += ', ' + v
line += ', ' + str(InstEval)

RankL.append(InstEval)
LineL.append(line)

return LineL, RankL

def calcSums(LineL, ADM = 4):

    RunAVGL = []
    IncVal = [0 for x in LineL[0].split(',')]  
    prvRank = 0
    topRank = 5
    NumFactors = 44
    Safetyclass = 0

    for i in range(len(LineL)):
        tmpR = LineL[i].split(',')
        r = [tmpR[0]] + [int(x) for x in tmpR[1:-1]] + [float(tmpR[-1])]

        currRank = r[-1]
        DCSM = sum(r[1:-1]) # (1000/45) * currRank
        RND = int(round(NumFactors * currRank / topRank, 0))
        MinND = RND - ADM
        MaxND = RND + ADM

        if (DCSM > MaxND):
            Safetyclass = 1
        elif (DCSM < MinND):
            Safetyclass = -1
        else:
            Safetyclass = 0

        LineL[i] += ', ' + str(Safetyclass)

        if (i == 0):
            RunAVGL.append(r + [DCSM])
        else:
            tmpL = [r[0]] + IncVal[1:]
            tmpIncVal = [''] + IncVal[1:]

            for x in range(1, len(r)):
                if x == len(r) - 1:
                    tmpL[x] = (r[x])
def main():
    term = []
    group = []
postfix = outDs

skip = 1
for c in CampusEvaluationFile:
    if skip == 1:
        skip = 0
        continue

    L = c.split(',
    campus.append(L[0])
campusEvaluation.append(L[:])
CampusEvaluationFile.close()

print("--- Reading terms and term groups .........")
term, group = getTermsAndGroups()
stats = open('DatasetStats' + postfix + '.csv', "w")
termLow, groupLow = termsToLower(term, group)

for ADM in tryADMs:
    for outClass in outClasses:
        if (outClass in [-1,0,1]):
            fileClass = str(outClass)
        else:
            fileClass = 'All'
            save = open('Dataset' + postfix + '_' + 'ADM' + str(ADM) + '_' + 'Class' + fileClass + '.' + OutputExt, "w")

            if (not WekaReady):
                writeFileHeader(save, term, group)

            print("--- Creating Dataset file: " + 'Dataset' + postfix + '_' + 'ADM' + str(ADM) + '_' + 'Class' + fileClass + '.' + OutputExt)

            LineL = []
            RankL = []
            LineL, RankL = createDataset(RankL, LineL, termLow, groupLow)
            LineL = sort(LineL, RankL)

            statistics = calcSums(LineL, ADM)

            if outClass in [-1,0,1]:
                LineL = selectClass(LineL, outClass)

            if (WekaReady):
                LineL = CleanForWeka(LineL)
writeDataset(save, LineL)
save.close()

writeDataset(stats, statistics)
stats.close()
print("--- Dataset is generated .......")

# Allows to Execute Code When the File Runs as a Script, but Not When It’s Imported as a Module
if __name__ == "__main__":
    main()