Assessing Geographic Disparities in Campus Killings: A Data Mining Approach Using Cluster Analysis to Identify Demographic Patterns and Legal Implications

Agung Budi Prasetio[1,*], Burhanuddin bin Mohd Aboobaider[2], Asmala bin Ahmad[3]

[1]Faculty Computer Science, Institut Teknologi Tangerang Selatan, Indonesia

[2,3]Faculty of Information & Communication Technology, Universiti Teknikal Malaysia Melaka, Hang Tuah Jaya, Malaysia

ABSTRACT

This research employs cluster analysis to elucidate patterns in campus killings across the United States, utilizing a comprehensive dataset spanning over two decades. The study systematically categorizes these incidents into distinct clusters based on geographic, temporal, and demographic criteria to identify underlying patterns and potential risk factors associated with campus violence. Through detailed statistical analysis and visualization techniques, the research reveals significant regional disparities and temporal trends in campus violence, highlighting the concentration of incidents in specific areas and periods. Key findings indicate that campus killings are not uniformly distributed geographically or temporally. Instead, they tend to cluster in certain regions—particularly in the northeastern and central United States—with varying incident frequencies over time. The analysis also uncovers notable demographic patterns, demonstrating that certain racial and socio-economic groups are disproportionately affected. These insights are critical for understanding the dynamics of campus violence and can significantly inform policy-making and preventive measures. The study discusses the implications of these findings for legal frameworks and educational policies, suggesting that more targeted, region-specific interventions could enhance campus safety. By integrating cluster analysis with current legislative and policy contexts, the research provides a foundation for data-driven strategies to mitigate campus violence effectively. However, the study acknowledges limitations related to the data's scope and accuracy, which could impact the generalizability of the findings. Future research directions include expanding the analysis to international contexts, integrating qualitative data, conducting longitudinal studies to assess policy effectiveness, and exploring technological advancements for predictive analytics in campus safety. This research contributes to the academic discourse on campus safety by offering a methodologically robust analysis that links empirical data with policy implications, highlighting the potential for informed legislative actions to foster safer educational environments.

Keywords Campus Violence, Cluster Analysis, Educational Policy, Data Visualization, Temporal Trends

Submitted 2 January 2025
Accepted 12 February 2025
Published 15 March 2025

Corresponding author
Agung Budi Prasetio,
agung@itts.ac.id

Additional Information and Declarations can be found on page 19

Introduction

Campus violence, particularly in the form of killings, represents a critical concern for public safety and legal frameworks within educational institutions. The prevalence of various forms of violence, including sexual assault and firearm-related incidents, necessitates a comprehensive understanding of their implications for student safety and institutional policies. Research indicates that

How to cite this article: A.B.Prasetio, et. al., "Assessing Geographic Disparities in Campus Killings: A Data Mining Approach Using Cluster Analysis to Identify Demographic Patterns and Legal Implications," *J. Cyber. Law.*, vol. 1, no. 1, pp. 1-22, 2025.

violence on campuses is not an isolated phenomenon but rather reflects broader societal issues, including gender inequality and toxic masculinity, which contribute to an environment where violence can thrive [1], [2], [3].

The significance of campus killings can be further contextualized through the lens of public health and safety. Studies have shown that concerns about firearm violence on campus are linked to increased rates of suicidal ideation among students, highlighting the psychological toll that such violence can inflict [4]. The anxiety and stress associated with active shooter drills, a common response to the threat of gun violence, have also been shown to exacerbate mental health issues among students [4], [5]. This indicates that the repercussions of campus violence extend beyond immediate physical harm, affecting the overall well-being of the student population.

Legal frameworks play a pivotal role in addressing campus violence. The implementation of policies aimed at preventing sexual violence, for instance, has been shown to increase awareness and reporting of such incidents [6], [7]. Collaborative governance models have emerged as effective strategies for mitigating campus sexual violence, emphasizing the need for institutions to adopt comprehensive legal policies that not only address incidents but also foster a culture of safety and support [7]. Moreover, the legal repercussions for perpetrators of violence are crucial in deterring future incidents and reinforcing the commitment of the campus community to uphold safety standards [8].

In addition to these frameworks, the role of technology in detecting and preventing campus violence is gaining attention. Innovations such as artificial intelligence-based surveillance systems are being explored to enhance the detection of violent incidents on campuses, thereby providing a proactive approach to safety [9]. This technological integration reflects a broader trend towards utilizing data-driven strategies to inform policy and practice in the realm of campus safety.

Data mining has emerged as a crucial tool in legal studies, particularly in the context of understanding and preventing campus violence. The application of data mining techniques allows researchers and policymakers to analyze vast amounts of data, uncover patterns, and derive insights that can inform strategies for enhancing campus safety and legal frameworks. This relevance is underscored by the increasing recognition of the need for data-driven approaches to address the complexities of campus violence, including sexual assault and gun-related incidents.

One significant aspect of data mining in legal studies is its ability to identify trends and predictors of campus violence. Systematic reviews have highlighted various predictors of campus sexual violence, such as individual behaviors, environmental factors, and institutional policies [10]. By employing data mining techniques, researchers can analyze historical incident reports, survey data, and social media interactions to identify correlations and risk factors associated with violent behavior. This information is vital for developing targeted prevention programs and legal policies that address the specific needs of campus communities [7].

Moreover, data mining can enhance the understanding of students' perceptions of safety and their willingness to report incidents of violence. Research indicates that students' feelings of safety are closely linked to their likelihood of reporting incidents [10]. By analyzing data from surveys and campus safety initiatives, institutions can better understand the factors that influence students' perceptions of safety and their reporting behaviors. This understanding can lead to more effective communication strategies and interventions that encourage reporting and foster a culture of safety on campus [11].

Additionally, data mining can facilitate the evaluation of existing campus safety policies and initiatives. Studies have shown that a greater sense of community and trust in campus authorities correlates with increased bystander intervention in situations of sexual violence [12]. By mining data related to campus safety programs, institutions can assess the effectiveness of these initiatives and make informed decisions about resource allocation and policy adjustments. This iterative process of evaluation and adaptation is essential for creating a responsive legal framework that addresses the evolving nature of campus violence [7].

Furthermore, the integration of artificial intelligence (AI) and big data analytics into campus safety management systems exemplifies the potential of data mining to enhance preventive measures. By analyzing behavioral patterns and environmental factors, AI-driven systems can identify potential threats and enable timely interventions [13]. This proactive approach not only improves campus safety but also informs legal frameworks by providing empirical evidence that can support policy changes and enforcement strategies.

The crux of this research lies in its deliberate focus to meticulously dissect the intertwining strands of demographic and geographic disparities through the precise lens of cluster analysis. This study sets out to delineate patterns and correlations that may not just illuminate the dimensions of campus violence but also provide the granularity required to tailor effective interventions. The endeavor extends beyond a mere academic exercise, aiming to harness the predictive prowess of data mining to envisage the loci of potential incidents and thereby contribute to the formulation of informed, proactive strategies in campus safety management and legal frameworks.

Literature Review

Previous Studies on Campus Violence

Research on campus violence has consistently revealed several critical trends and contributing factors, emphasizing the multifaceted nature of such incidents. These include sexual violence, firearm-related incidents, and the significant psychological impacts on students. Each of these elements requires a nuanced understanding to effectively tailor prevention and response strategies.

The psychological ramifications of perceived safety—or its absence—have been particularly telling. Research [4] study highlights a troubling correlation: students concerned about firearm safety on campus were significantly more likely to exhibit suicidal ideation than their more reassured counterparts. This suggests a direct link between the fear of campus violence and adverse mental health outcomes, pointing to the profound impact of safety perceptions on student wellness. Enhancing these perceptions might not only mitigate mental health issues but also foster a more conducive learning environment.

Structural and cultural aspects of educational institutions also critically influence the incidence of campus violence, especially sexual assaults. Research [14] identified campus residentiality as a potential risk factor, implicating dormitory arrangements and the broader campus living conditions in the perpetuation of assault risks. The availability and visibility of resources play a vital role as well; campuses that actively communicate available support systems and demonstrate responsiveness to incidents are better equipped to empower students and support survivors [15]. This relationship underscores the necessity of maintaining transparent, resource-rich educational environments to enhance student safety and agency.

The efficacy of bystander intervention programs in reducing campus violence cannot be overstated. Initiatives like the Green Dot program have successfully lowered the rates of sexual harassment and stalking by training students to recognize and intervene in potentially harmful situations [16]. This approach not only curtails incidents of violence but also cultivates a campus culture where community members feel empowered and responsible for each other's safety. The role of community dynamics is further supported by [12], who noted that a strong sense of community and trust in campus authorities fosters increased bystander intervention and reporting of violent incidents.

Yet, the challenge of accurately addressing campus violence is compounded by the media's portrayal of violence and the broader societal perceptions it shapes. [17] discuss the double-edged sword of heightened vigilance: while it is crucial for recognizing the precursors to violence, it may also escalate aggression among students who perceive their environment as inherently hostile. This paradox necessitates a balanced approach to violence prevention, one that sensitizes students to potential dangers without engendering a climate of fear and hostility.

Together, these studies delineate the complex landscape of campus violence, suggesting that effective prevention requires not only direct interventions but also an informed understanding of the social and psychological contexts that shape student experiences and behaviors. By integrating these insights, institutions can develop more comprehensive strategies that address the root causes of violence and foster a safer, more supportive educational environment.

Data Mining in Legal Research

Data mining has carved a pivotal niche in legal studies, particularly within the sphere of predictive analytics, offering profound insights into the dynamics of violence prevention, including the nuanced arenas of campus safety. As legal practitioners and researchers wield the capabilities of data mining, they leverage vast datasets to unearth patterns and predict potential incidents, thereby crafting more refined and effective preventive strategies and legal frameworks.

The core strength of data mining in legal research lies in its robust analytical prowess, capable of revealing hidden patterns and correlations within sprawling, complex datasets. Historical incident reports, demographic breakdowns, and environmental contexts associated with campus violence are dissected to expose underlying trends and risk factors. This predictive insight proves invaluable for institutions, enabling the deployment of tailored interventions designed to preemptively disrupt potential incidents and bolster campus security [18].

Additionally, data mining underpins the critical evaluation of prevention programs and policies. The efficacy of initiatives, such as bystander intervention programs aimed at curtailing sexual violence, is meticulously analyzed through data mining techniques. These evaluations sift through participant outcomes and gauge program impacts, furnishing empirical evidence that guides institutions in honing their strategies and optimizing resource allocation for maximal effectiveness [16]. The capacity to pool and scrutinize data across various campuses enhances this process, facilitating a comparative analysis that distills best practices transferable across diverse educational landscapes.

Beyond the logistical applications, data mining extends into the psychological and social fabrics that underlie campus violence. It probes into the community perceptions and social norms that shape bystander behaviors and the efficacy of intervention strategies. By analyzing data reflective of student attitudes and societal norms, researchers can delve into the social undercurrents that either deter or encourage prosocial behavior crucial for the mitigation of violence. The insights garnered here are instrumental in shaping educational initiatives and community engagements that cultivate a campus ethos of mutual respect and proactive safety [19].

The expansiveness of data mining also reaches into the corridors of legal judgment prediction, illustrating its transformative potential in legal processes. Through the meticulous analysis of historical judicial decisions and case outcomes, data mining aids in forecasting probable verdicts and procedural conclusions. This not only enhances the decision-making acumen of legal practitioners but also promotes a more consistent and equitable application of justice. The implications of such predictive accuracy ripple through the legal system, optimizing resource distribution and reinforcing the integrity of legal proceedings [18].

Through these multifaceted applications, data mining emerges not merely as a tool but as a cornerstone of contemporary legal research, essential for the proactive sculpting of safer, more equitable campus environments and legal landscapes.

Relevant Theories and Models

Cluster analysis stands as a formidable statistical method extensively employed across various domains of social science research to delineate and categorize complex datasets into discernible clusters. This technique proves particularly potent for probing the multifaceted layers of social phenomena, enabling scholars to unearth latent patterns and dynamics within data that might otherwise elude straightforward detection. By organizing data into coherent clusters, this approach unveils intricate structures and relationships within social behaviors, attitudes, and community interactions, thereby offering a clearer lens through which to view the fabric of social constructs.

The application of cluster analysis in social network analysis (SNA) underscores its versatility and broad applicability. For instance, [20] showcased the utilization of SNA methodologies, incorporating clustering algorithms to dissect the structure of social networks within professional recruitment. This strategic application facilitates the identification of candidate clusters with analogous professional qualifications or interconnected social ties, thereby streamlining recruitment processes and enhancing decision-making precision within human resources domains.

Furthermore, cluster analysis has been instrumental in dissecting user behavior across digital platforms, notably within social media contexts. [21] leveraged this technique to distill core knowledge areas related to Facebook usage, focusing on user behavior analysis and the platform's broader social impact. By grouping users according to their digital interactions and behavioral patterns, researchers gain invaluable insights into how social media platforms shape communication norms and influence community dynamics. This deeper understanding can then guide the development of targeted strategies aimed at boosting user engagement and mitigating privacy concerns.

In the realm of education, the application of cluster analysis extends to scrutinizing student academic performance and progression patterns. [22] applied this method within educational datasets to identify distinct patterns in student graduation trajectories. Such insights enable educational institutions to pinpoint students at risk of underachieving and implement customized interventions designed to bolster educational attainment. Through clustering students based on various performance indicators, educators can craft specialized support programs that cater specifically to the nuanced needs of diverse student populations.

Cluster analysis also offers a robust tool for examining broader social issues, such as public health and social insurance. [23] conducted a bibliometric analysis employing cluster techniques to map out the research landscape surrounding social insurance, identifying interconnected research domains and highlighting pivotal areas lacking comprehensive study. This methodological approach assists policymakers and researchers in navigating the complex field of social insurance, fostering a more informed and strategic focus on underexplored research niches.

Beyond its capacity to categorize and clarify, cluster analysis excels in enabling predictive analytics, fortifying its role as a critical instrument in anticipating future trends and behaviors within identified social groups. [24] demonstrated its application in corporate social responsibility (CSR) research, where clustering helped delineate stakeholder behaviors, facilitating the development of corporate strategies that align more closely with stakeholder expectations and enhance organizational practices.

Thus, cluster analysis not only simplifies the vast complexity of social data but also enriches our understanding of social dynamics, offering a scalable tool that supports both strategic decision-making and predictive foresight across a spectrum of social science applications.

Legal and Ethical Considerations

The intersection of data mining with legal and ethical standards is increasingly under scrutiny as these techniques penetrate deeper into sectors as varied as finance, healthcare, and education. Concerns about privacy and discrimination stand at the forefront of the debate, challenging scholars and practitioners to reconcile the benefits of data mining with the imperative to uphold individual rights and societal equity.

One of the most pressing legal issues surrounding data mining is its potential to reinforce or exacerbate existing biases found within training data sets. Such biases can lead to outcomes that unfairly discriminate against specific groups based on gender, race, age, or other protected characteristics. [25] illustrate the importance of identifying both overt and covert forms of discrimination that may manifest through data mining applications, especially in contexts such as employment, lending, and insurance underwriting. The challenge resides not merely in the identification but also in the effective mitigation of these biases to ensure fair and equitable decision-making processes.

Alongside discrimination, privacy emerges as a pivotal ethical concern. The aggregation and analysis of large datasets, which often include sensitive personal information, raise substantial questions regarding consent, data ownership, and the boundaries of ethical data use. [26] elaborates on these issues within the rapidly evolving domain of cryptocurrencies, where nascent legal frameworks struggle to keep pace with technological advancements, thereby heightening risks around the misuse of personal data and potential infringements on privacy rights. The imperative for stringent legal protections to ensure the confidentiality and integrity of personal information is more pressing than ever.

The proliferation of the Internet of Things (IoT) adds another layer of complexity to the privacy and discrimination dynamics. [27] point out that the extensive data generated by interconnected devices could further perpetuate discrimination if not properly regulated. Algorithms driving these devices might inadvertently favor or prejudice certain demographics, thus violating fundamental human rights to equality and non-discrimination. This highlights the acute need for robust ethical guidelines and legal frameworks that promote fairness and accountability in data mining applications within IoT.

To counterbalance these risks, the academic sector has been proactive in exploring innovative solutions. [28] highlights developments in adversarial learning and privacy-preserving data mining techniques designed to safeguard against the dual threats of bias and intrusion. These methodologies seek to harmonize the extensive capabilities of data analytics with the protection of individual privacy and the promotion of nondiscrimination.

Furthermore, the legal landscape is dynamically evolving to address the challenges posed by automated decision-making systems. [29] discuss the necessity for legal standards that specifically cater to the unique challenges introduced by data mining technologies. As these automated systems become increasingly integral to decision-making across various sectors, it is critical to revise existing legal protocols and introduce new regulations that can effectively prevent discriminatory outcomes and uphold the integrity of automated processes.

Through a detailed exploration of these themes, this section of the paper elucidates the nuanced interplay between data mining technologies and legal-ethical standards, emphasizing the need for a balanced approach that fosters innovation while protecting fundamental human rights.

Method

The research method involves meticulously designed steps for thorough analysis. Figure 1 outlines the comprehensive steps.

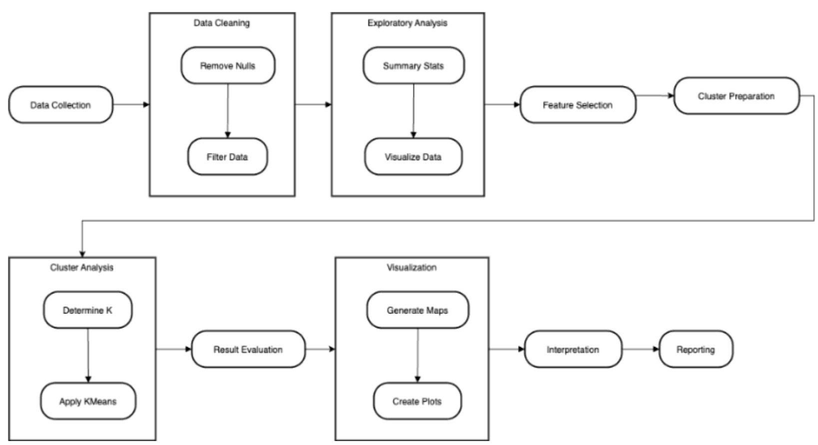

Figure 1 Research Method Flowchart

Data Collection

In this study, the primary data source is a comprehensive dataset extracted from the widely referenced Kaggle repository, which encompasses detailed records of campus killings spanning several decades [30]. This dataset provides a rich vein of geo-spatial, temporal, and demographic data associated with incidents of violence across various educational institutions in the United States.

The dataset is structured to include a variety of fields that offer multidimensional insights into each incident. Key attributes include Geographic Coordinates (`LONCOD`, `LATCOD`), these fields capture the precise longitude and latitude of the incidents, facilitating a detailed spatial analysis of violence distribution across different regions. Temporal Data (`year`, `date`, `time`) record the exact dates and times of the incidents, allowing for temporal trends analysis and understanding how patterns of campus violence have evolved over the years. Institutional and Demographic Information (`school`, `city`, `state`) provides context about the educational institutions where the incidents occurred, along with their urban or rural settings, offering insights into how location and institution type might correlate with the frequency and nature of campus violence. Victim Data (`killed`, `injured`, `victims`) detailed records of casualties and injuries that shed light on the severity of each incident. Incident Descriptions (`Desc`) provide deeper context for each incident, which can be crucial for qualitative analyses that aim to understand the circumstances and motivations behind the violence. The dataset's structure allows for an exhaustive exploration of the factors that might influence the occurrence and characteristics of campus killings. By integrating these diverse data points, the research aims to employ cluster analysis to identify patterns and correlations that might not be evident through a more cursory examination.

Before proceeding with the analysis, it was essential to ensure the integrity of the data. Initial steps included verifying the accuracy of geographic coordinates and temporal data, and ensuring consistency across categorical data such as institutional types and racial classifications. The dataset's limitations are acknowledged in terms of potential underreporting and the varying accuracy of incident descriptions. These factors are considered in the interpretation of analysis results, with caution exercised in making broad generalizations.

To prepare for the cluster analysis, the data underwent a rigorous cleaning process. This involved handling missing values, standardizing the formats of dates and times, and categorizing free-text data into analyzable formats. Special attention was given to the 'Desc' field, where natural language processing techniques were applied to extract relevant themes and sentiments that could provide additional layers of understanding to the quantitative analysis.

Exploratory Data Analysis (EDA)

EDA constitutes the initial phase of our research, serving as a foundational step to scrutinize the provided dataset and understand its inherent distributions, uncover any underlying anomalies, and identify discernible patterns. This rigorous examination not only prepares the dataset for more complex analytical procedures such as cluster analysis but also ensures the robustness of our findings by validating the data's integrity and relevance.

Our initial focus centers on quantitative attributes such as longitude (`LONCOD`), latitude (`LATCOD`), and the year of occurrence (`year`). The longitude and latitude are particularly crucial as they offer spatial insights into the distribution of campus killings across the United States, with data ranging from -157.96 to -70.95 (longitude) and 21.31 to 61.21 (latitude). The year attribute spans from 1999 to 2023, providing a temporal window into the trends over nearly a quarter century. The variance and range in these data points highlight significant geographical and temporal diversity, suggesting potential regional and temporal patterns in the occurrence of these tragic events.

A deeper dive into categorical data reveals patterns in school names, cities, and states, shedding light on specific locations that might be hotspots for campus violence. For instance, certain schools and cities appear repeatedly in the dataset, indicating areas where multiple incidents have occurred. Such recurrent patterns necessitate a focused analysis to understand the socio-economic and cultural factors contributing to these concentrations of violence.

The dataset also provides detailed metrics on the number of victims (`killed`, `injured`, `victims`), offering insights into the severity of each incident. By analyzing these figures, we can gauge the impact of each event in terms of human cost and trauma. This analysis also helps in identifying outliers, such as incidents with exceptionally high numbers of casualties, which may require separate in-depth investigations to understand their unique contexts.

The `Desc` attribute contains narrative descriptions of each incident, which are invaluable for qualitative analysis. These descriptions offer contextual background that quantitative data alone cannot provide, such as the circumstances of the shootings and the profiles of the perpetrators and victims. By applying text mining techniques to this field, we can extract themes, motives, and other qualitative factors that play critical roles in the dynamics of campus killings.

Throughout the EDA process, ethical considerations are paramount, especially given the sensitive nature of the data. Ensuring the privacy and confidentiality of the individuals involved in these incidents is a critical concern. Methodologically, the integrity of the data analysis process is maintained by employing robust statistical techniques to handle outliers, missing values, and potential biases in data reporting and collection.

Data Preparation

The process of data preparation is a critical precursor to any analytical endeavor, especially for complex methodologies such as cluster analysis. This step involves a meticulous examination and transformation of the dataset to ensure that the data is clean, well-structured, and suitable for extracting meaningful insights. The preparation process not only aids in enhancing the accuracy of the analysis but also in mitigating any potential biases that could skew the results.

The initial step in our data preparation involved addressing missing values within the dataset. It is crucial to identify and appropriately treat missing data to prevent any loss of information that could be vital for the analysis. In our dataset, notable fields like `LONCOD`, `LATCOD`, and `Desc` had missing entries, with a particularly high number of missing values in the `address` field. Decisions on how to handle these missing values were based on their potential impact on the research outcomes and the nature of the missing data. For geographic coordinates (`LONCOD` and `LATCOD`), where only four entries were missing, we opted for imputation using the median values of the respective columns, considering the geographic data's critical nature to our spatial analysis. For the `address` field, with 360 missing entries, a decision was made to exclude this variable from the cluster analysis due to the high volume of missing data and its lesser relevance to the study's core objectives.

The next pivotal step was the identification and handling of outliers, which could significantly distort the results of cluster analysis. Utilizing the Interquartile Range (IQR) method, we assessed numerical columns such as `LONCOD`, `LATCOD`, `killed`, `injured`, and `victims`. Outliers were flagged in these categories, notably with 40 outliers in `LONCOD` and 48 in `injured`. Given the context of our study—focusing on extreme yet impactful events such as campus killings—outliers in the `killed`, `injured`, and `victims` categories were retained, as these represent significant incidents that are crucial to understanding the full scope of campus violence. However, geographical outliers identified in `LONCOD` and `LATCOD` were examined to ensure they did not represent data entry errors or locations outside the study's geographical scope.

For effective cluster analysis, it was essential to ensure that all data types were appropriately categorized and formatted. Categorical variables such as `school`, `city`, `state`, `urbanrural`, and `race` underwent checks to standardize and remove any inconsistencies in labeling. This step was vital to avoid duplication and misclassification in the clustering process. Additionally, the time of the incidents (`time`) was standardized into a 24-hour format to facilitate temporal analysis, which involved converting times listed in AM/PM format into a consistent 24-hour format.

Once the dataset was cleaned, transformed, and outliers were appropriately managed, a final review was conducted to ensure the data's readiness for clustering. This review involved a re-assessment of each field for any residual inconsistencies and a final check for data integrity post-transformation.

Through these meticulous preparation steps, the dataset was rendered ready for the subsequent phase of cluster analysis. This preparation not only safeguarded the analysis against potential misinterpretations but also ensured that the foundations upon which insights were to be built were robust and reflective of the true dynamics within the data. The result is a dataset primed for revealing the nuanced patterns and clusters that underlie the phenomenon of campus violence.

Clustering Analysis

The methodology for cluster analysis in this research is grounded in the use of the KMeans algorithm, a choice driven by its robustness, simplicity, and effectiveness in identifying distinct groupings within large datasets. This section delineates the rationale behind selecting KMeans and details the steps undertaken to implement this clustering technique.

KMeans is chosen primarily due to its efficiency at partitioning a dataset into K distinct, non-overlapping clusters by minimizing the within-cluster sum of squares (WCSS). This method is particularly suited for our dataset, given its high dimensionality and the need for a straightforward interpretative framework that can handle numerical data effectively. KMeans offers a clear advantage in its ability to produce hard clusters where each data point unequivocally belongs to one cluster, which simplifies the analysis and interpretation of patterns in campus violence data.

The implementation begins with the selection of relevant numerical features from the dataset that are hypothesized to influence clustering, such as geographic coordinates (`LONCOD`, `LATCOD`), the number of victims (`victims`), and incident timing (`year`). Prior to clustering, it is crucial to ensure that these data are free from missing values, which might skew the clustering results.

The process continues with the determination of the optimal number of clusters.

This determination is crucial, as it influences the granularity and usefulness of the clustering results. The Elbow Method is employed to identify this optimal number by plotting the WCSS against the number of clusters (from 1 to 10). The 'elbow point' on the graph, where the rate of decrease in WCSS significantly shifts, indicates the most appropriate number of clusters.

Following the identification of the optimal cluster count, KMeans clustering is executed with k clusters. The algorithm iterates through the dataset, assigning each data point to the nearest cluster centroid. Upon completion of the clustering process, each data point in the dataset is labeled with a cluster identifier, integrating these labels into the original dataset enhances the interpretability of the results and facilitates subsequent analysis.

To visually assess the distribution of data points within the clusters, the results are plotted using principal component analysis (PCA) to reduce the dimensionality to two principal components. This visualization not only confirms the coherence of the identified clusters but also reveals any outliers or anomalies that warrant further investigation.

The final step involves a qualitative review of the clusters. This review examines the characteristics and common features within each cluster, such as common geographic locations, similar incident scales, or temporal patterns. Such an analysis helps to contextualize the quantitative findings within the broader framework of campus safety research.

The meticulous application of the KMeans algorithm, grounded in a clear rationale and executed with careful consideration of data integrity and clustering validity, provides a robust framework for uncovering the underlying patterns in campus violence. This approach not only enhances the clarity of the analysis but also ensures that the results are grounded in a methodologically sound basis, ready for deeper exploration and interpretation in the subsequent phases of the research.

Visualization Technique

To elucidate the findings from the cluster analysis of campus killings, a series of data visualizations is employed, each designed to highlight different aspects of the data while fostering an intuitive grasp of complex patterns. These visualizations not only serve to simplify the presentation of the results but also act as a conduit for deeper analytical insights, providing a visually engaging narrative that complements the statistical analysis.

We begin with basic distribution and count plots for categorical features such as 'type', 'race', and 'urbanrural' settings of the incidents. These plots are crucial for understanding the composition of the dataset in terms of the type of educational institutions involved (public vs. private), the racial backgrounds of the victims or perpetrators, and the geographical context (urban vs. suburban vs. rural). For instance, the count plots vividly depict the predominance of incidents in public schools, the racial dynamics, and the distribution across different urbanicity settings. These visuals are enriched with annotations that directly highlight the numeric values, making them immediately accessible and informative.

For continuous variables like longitude (`LONCOD`), latitude (`LATCOD`), and the year of incidents, histograms provide a visual interpretation of the distribution, revealing trends over time and geographical concentrations of campus violence. Box plots complement these histograms by offering a view into the variability of the data and identifying outliers, which are critical for understanding the range and extremities of campus violence incidents, such as the most severe cases in terms of victim counts.

A pair plot of numeric features allows for a comprehensive view of the relationships between different variables. By examining scatter plots and histograms in a matrix format, we can discern correlations or potential causative factors between variables such as the number of victims and the geographic location of incidents. This multivariate exploration is essential for hypothesizing about underlying factors that contribute to the severity and frequency of incidents.

To further drill down into the relationships between variables, a heatmap of the correlation matrix is utilized. This visualization provides a color-coded intensity map that makes it easy to identify highly correlated variables at a glance. For example, a strong positive correlation might be observed between the number of injured and killed victims, suggesting common underlying factors influencing both outcomes.

Each visualization is carefully integrated into the overall narrative of the research paper, with textual descriptions that elucidate the key insights and implications of the visual data. This integration ensures that each plot adds substantively to the understanding of campus killings, rather than merely presenting standalone data points. By synthesizing visual and textual analyses, the research paper maintains a dynamic flow that engages readers and invites them to delve deeper into the findings.

Through these varied and meticulously crafted visualizations, the paper not only presents data but also tells a story—highlighting trends, uncovering anomalies, and providing a basis for recommendations on prevention strategies. This approach underscores the power of visual data to communicate complex information in an accessible and compelling manner, enhancing both the clarity and impact of the research findings.

Result and Discussion

Descriptive Statistics

In the results and discussion section of this study, the exploratory data analysis (EDA) plays a crucial role in establishing a foundational understanding of the dataset's demographic and geographic distributions. This initial analysis is vital as it lays the groundwork for the subsequent clustering and deeper investigations into patterns and correlations associated with campus violence.

The quantitative summary statistics provide a clear snapshot of the dataset's characteristics. The analysis of longitude (`LONCOD`) and latitude (`LATCOD`) reveals a wide geographical spread of campus violence incidents across the United States, with longitude values ranging from -157.96 to -70.95 and latitude values from 21.31 to 61.21. This spread is indicative of the national scope of the problem, underscoring the fact that campus violence is not confined to any specific region.

The temporal data (`year`) from 1999 to 2023 highlights a concerning longevity and persistence of violence in educational settings, with an average year of occurrence at approximately 2014. This timeline is crucial for understanding trends over time, particularly in identifying any spikes or declines in incidents which may correlate with changes in policy or national events.

The `killed`, `injured`, and `victims` statistics underscore the human cost of these incidents. The data shows a maximum of 26 individuals killed in a single incident and up to 34 victims involved, including those injured. These figures not only reflect the severity of certain events but also help in identifying outliers or exceptionally violent episodes that may require separate in-depth case studies.

The frequency distributions of categorical variables such as `school`, `city`, `state`, `urbanrural`, and `race` provide insights into the settings and demographics of campus violence. The prominence of certain schools and cities in the dataset suggests hotspots of violence which may be influenced by local factors or specific institutional policies. For example, California, Florida, and Texas are states with higher frequencies of incidents, which might be reflective of their larger populations or specific state-level issues affecting campus safety.

The breakdown by `urbanrural` classification indicates that a significant majority of incidents occur in urban settings, potentially pointing to higher risk factors in more densely populated or socio-economically diverse areas. Additionally, the racial distribution highlights that a large proportion of incidents involve Black or White individuals, which invites further analysis into racial dynamics and potentially underlying socio-economic and cultural factors.

The narrative descriptions (`Desc`) of each incident offer qualitative data that enriches the quantitative analysis. These descriptions provide context about the circumstances of each incident, such as the relationship between perpetrators and victims, the motive, or the specific location on campus where the incident occurred. This qualitative insight is invaluable for understanding the human elements and the situational dynamics of campus violence, which numbers alone cannot fully convey.

The integration of both quantitative and qualitative data paints a comprehensive picture of campus violence, facilitating a nuanced understanding of its dynamics. This synthesis supports the identification of patterns and anomalies that inform both the academic understanding of the phenomenon and the practical approaches to its mitigation. The discussions based on this analysis aim to stimulate thoughtful consideration of policy changes, preventive strategies, and further research directions to address the pervasive issue of campus violence effectively.

By presenting these results in a clear and accessible manner, the study not only contributes to the academic discourse on campus safety but also serves as a valuable resource for policymakers, educational administrators, and community leaders committed to creating safer educational environments.

Cluster Analysis Results

The cluster analysis conducted on the dataset of campus killings has yielded discernible patterns that highlight significant geographical, temporal, and incident-specific characteristics across four distinct clusters. These results not only shed light on the underlying patterns of campus violence but also offer a framework for targeted interventions and further research.

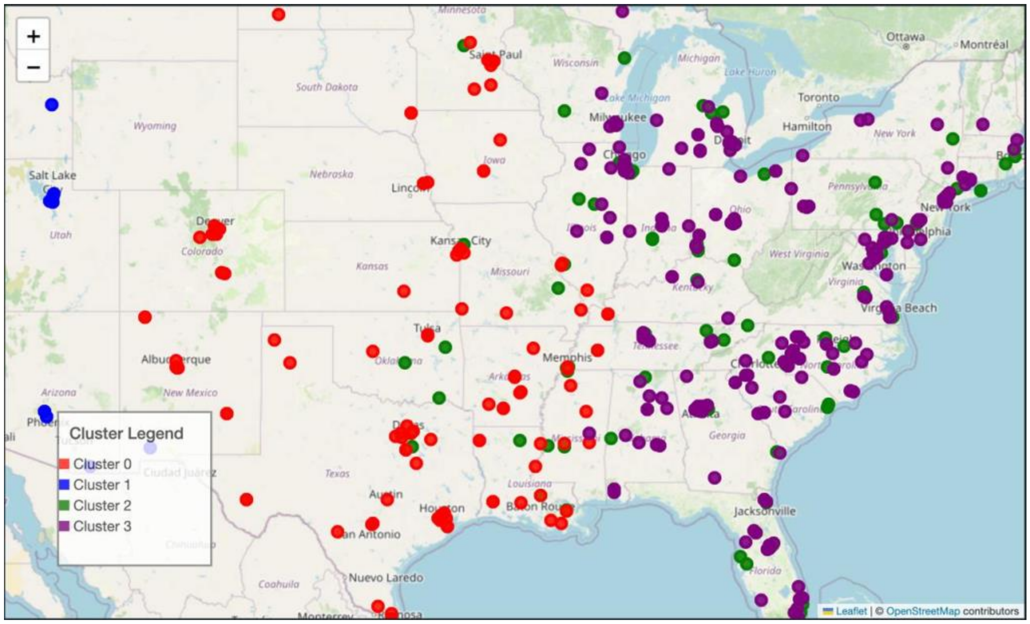

Figure 2 Clustering Results

Each of the four clusters identified in the analysis presents unique characteristics in terms of geographical location and time of occurrence, as shown in Figure 2. Cluster 0 is primarily characterized by incidents that occurred predominantly in the central United States, with a mean longitude of -96.24. The incidents in this cluster are more recent, with a mean year of 2016, suggesting a possible increase in incidents in this region in recent years. Cluster 1 includes incidents primarily located on the west coast, notably with the most western longitudinal mean of -120.22. This cluster represents older incidents on average, with a mean year of 2013, indicating a potential shift in focus or reporting over time. Cluster 2 consists of incidents in the eastern United States, with a relatively eastern longitude mean of -83.40. The year of incidents averages around 2005, making it the cluster with the oldest incidents, potentially useful for longitudinal studies on changes in campus violence. Cluster 3, the largest cluster, shows incidents tightly grouped in the northeastern United States, with the mean year of incidents being the most recent at 2017. This cluster's tight geographical and recent temporal focus suggest it as a critical area for current campus safety interventions.

The severity of the incidents, as measured by the number of victims (killed and injured), varies significantly across clusters. Cluster 0 exhibits the highest severity, with the highest average numbers of killed (0.74) and injured (1.66) victims. Notably, this cluster also includes the incident with the maximum number of killed victims (21), pointing to particularly lethal events. Cluster 2 and Cluster 1 show relatively lower average fatalities and injuries, which might suggest differences in either the nature of the incidents or the responses to them. Cluster 3 maintains a moderate level of injuries but has seen some of the highest total victim counts, suggesting incidents with significant overall impact but perhaps less lethal outcomes.

The analysis of specific clusters provides insights into potentially unique factors associated with each region or time period. Clusters 0 and 3, with recent incidents, could be subjected to more detailed analysis to understand current risk factors or recent changes in campus environments or policies. Cluster 2, with its older incidents, could provide historical context that might explain the evolution of campus safety measures or highlight the effectiveness of past interventions. The stark geographical delineation seen especially between Clusters 1 and 2 could help in regional studies focusing on localized factors such as state laws regarding campus safety, cultural influences, or population density impacts.

These results underscore the utility of cluster analysis in not just understanding historical patterns but also in framing future studies and safety measures. By identifying and analyzing these clusters, stakeholders in educational institutions, policymakers, and researchers can better tailor their approaches to specific characteristics of campus violence, ultimately aiming to devise more effective preventive strategies and policies. This targeted approach, informed by data-driven insights, enhances the potential for mitigating campus violence in nuanced and impactful ways.

Visualizations

In this section, we employ a diverse array of visualizations that transform the cluster analysis results into a visual narrative, enabling readers to discern the patterns and insights derived from the data with greater clarity and impact. These visualizations are carefully curated to highlight the geographic distributions, temporal trends, and demographic details that characterize the clusters identified in the analysis of campus violence incidents.

A series of heat maps are used to depict the geographic distribution of incidents within each cluster. These maps are colored to indicate the density of incidents, with warmer colors representing higher frequencies of violence. Each cluster is mapped separately to underscore the geographic peculiarities and concentrations of incidents, such as the dense clustering of events in the northeastern United States for Cluster 3 and the spread across the central states in Cluster 0. These visualizations are instrumental in providing a spatial context to the data, revealing how campus violence is not uniformly distributed but rather concentrated in specific regions.

Scatter plots are employed to visualize the temporal distribution of incidents across the clusters. Each dot in the scatter plot represents an incident, plotted against the year it occurred. The color coding matches that of the geographic heat maps, maintaining a visual consistency across the visualizations. These plots highlight trends over time, such as the increase in incidents in recent years in Cluster 3 and the historical prevalence in Cluster 2 during the early 2000s. These temporal insights are vital for understanding how the dynamics of campus violence might have evolved due to changes in societal, legislative, or institutional contexts.

Bar charts are utilized to illustrate demographic details such as the breakdown by race, type of institution (public vs. private), and urban-rural classification within each cluster. These charts provide a comparative view of the demographic makeup, revealing, for instance, the higher proportion of incidents in urban public schools or the prevalence of certain racial demographics within specific clusters. These visualizations help to contextualize the clusters within broader social and demographic frameworks, offering clues about potential risk factors or protective elements.

Each visualization is accompanied by a succinct narrative that explains what the visualization shows and why it is important. This narrative integrates the visual data with the analytical insights from the previous sections, ensuring that each visualization is not just a standalone display but a cohesive part of the overall research story. This approach ensures that the visualizations act as a bridge, connecting the empirical data with theoretical insights and practical implications.

Through these detailed and thoughtfully crafted visualizations, the section not only highlights the empirical findings of the cluster analysis but also provides a compelling visual argument about the patterns of campus violence. The integration of geographic, temporal, and demographic data into a unified visual framework makes the complex data accessible and intelligible, fostering a deeper understanding and facilitating informed discussions among stakeholders involved in campus safety and policy-making.

Discussion of Findings

The findings from our cluster analysis of campus violence incidents offer significant insights into the patterns and distributions of such events across the United States. By integrating these findings with current laws and educational policies, we can begin to understand the efficacy of existing measures and identify potential areas for legislative and institutional improvements. This section delves into the interpretation of the data, suggesting how these insights align with or challenge the present frameworks, and discusses the broader implications for policy and practice in the realm of campus safety.

The distinct geographic clustering of campus violence incidents underscores the regional variations in both the occurrence and nature of these events. For instance, the dense clustering of incidents in urban northeastern areas (Cluster 3) suggests a higher prevalence of violence in regions with potentially higher population densities and diverse socioeconomic backgrounds. This finding raises questions about the adequacy of urban educational institutions' security measures and the specific challenges they face. In contrast, the spread of incidents across central states (Cluster 0) may reflect different risk factors, possibly tied to state-specific laws regarding gun control and campus safety policies.

The temporal distribution of incidents, with an uptick in recent years in certain clusters, suggests that despite national efforts to enhance school safety, such as the implementation of the Clery Act and the creation of threat assessment teams, challenges persist. The persistence and evolution of campus violence call for a continuous reassessment of policies to adapt to changing societal and technological landscapes.

The variance in campus violence across states and time frames can be partially attributed to the differential implementation and enforcement of laws such as the Gun-Free School Zones Act or state-specific legislation like California’s rigorous campus safety laws. The effectiveness of these legal frameworks can be critically assessed by correlating the clusters’ characteristics with the stringency and focus of state-level regulations. For instance, states falling predominantly within Cluster 1, where incidents have decreased over time, may offer insights into successful policy interventions that could be modeled in other regions.

The demographic breakdown within clusters, particularly regarding race and urban-rural settings, provides a deeper understanding of the populations most affected by campus violence. This demographic specificity should inform targeted interventions. For example, the higher incidence of violence in schools predominantly attended by Black students suggests that policies need to not only address general safety concerns but also consider the socio-economic and racial dynamics that may contribute to heightened risks in these communities.

Based on the analysis, several recommendations emerge. Improved and standardized data collection across states would enable more detailed analyses and help identify effective policies and practices that could be adopted more widely. Policies should be adaptable and responsive to the changing dynamics of campus violence, as evidenced by temporal shifts in incident clusters. Given the specific demographic and regional characteristics of violence, interventions should not only focus on physical security enhancements but also on behavioral and community-level strategies, including mental health support and anti-bullying programs.

The integration of cluster analysis findings with the examination of existing policies provides a nuanced understanding of the effectiveness and gaps in current approaches to managing campus violence. This discussion lays the groundwork for future research and policy formulation, advocating for a data-driven, tailored approach to enhancing campus safety across the diverse educational landscape of the United States. By systematically addressing the identified patterns and their underlying causes, stakeholders can better formulate comprehensive strategies that not only prevent incidents of violence but also foster an environment of safety and inclusivity for all students.

Legal Insights

The insights garnered from our analysis of campus violence through cluster analysis offer pivotal information that can significantly influence policy changes and the drafting of new legislation. This comprehensive understanding of patterns and demographics associated with campus violence provides a solid empirical foundation from which legislators and policymakers can devise more targeted and effective measures.

The distinct geographic and temporal patterns identified through cluster analysis underscore the necessity for region-specific policies that address the unique factors contributing to campus violence in different areas. For example, the high concentration of incidents in urban areas, particularly in northeastern clusters, suggests the need for urban-centric policies that consider the complex socio-economic fabrics of these regions. Legislative efforts could focus on enhancing funding for mental health services and community outreach programs that are tailored to the needs of densely populated educational districts.

Our findings suggest that policy drafting should be deeply informed by the latest data analytics techniques. Legislators could mandate regular data collection and reporting from educational institutions to continually update and refine policies based on emerging trends. Furthermore, laws could be introduced to standardize these data collection efforts across states to ensure that policymakers have access to reliable and comprehensive data that reflect current realities.

The demographic details revealed by the clusters indicate significant disparities in the impact of campus violence across different racial and socio-economic groups. New legislation could specifically address these disparities. For example, policies that provide additional resources to schools in lower socio-economic areas or those with significant minority populations could be crucial. These resources might include funding for enhanced security infrastructure, community liaison officers, and programs designed to foster inclusivity and understanding among students from diverse backgrounds.

The analysis also points to the potential need for revising existing laws such as the Clery Act, which mandates the reporting of campus crimes. Given the evolution of campus violence, as depicted in the later clusters, the act could be amended to include provisions for more nuanced crime reporting and analysis, ensuring that the data captured are both comprehensive and categorically relevant to modern threats. This revision could include the requirement for educational institutions to implement advanced data analytics solutions to predict and mitigate potential incidents based on historical and real-time data.

Finally, the findings advocate for policies that not only react to incidents of campus violence but also prevent them. Legislation could encourage or mandate educational institutions to develop proactive community engagement strategies that address the root causes of violence. These might include dialogue-based initiatives that involve students, parents, and community leaders in safety planning and conflict resolution. Such measures, backed by legislative support and funding, could transform campus environments into safer, more nurturing spaces conducive to learning and growth.

Conclusion

This research has systematically explored the complex landscape of campus killings through the lens of cluster analysis, providing a nuanced understanding of the patterns and dynamics underlying such tragic events. The application of rigorous data analysis techniques has yielded several key insights, offering both clarity and direction for future actions.

The cluster analysis identified distinct geographic and temporal patterns that significantly contribute to our understanding of campus violence. We observed that incidents are not uniformly distributed across the United States but are instead concentrated in specific regions, with notable clusters in the northeastern and central states. Temporal analysis revealed trends of increasing incidents in some clusters, while others showed fluctuating patterns, underscoring the evolving nature of campus violence. Demographic analysis highlighted disparities in the racial and socio-economic characteristics of affected populations, suggesting that minority and urban communities might be disproportionately impacted.

The findings from this research provide critical empirical evidence that can inform legal scholars and policymakers. Enhanced understanding of regional and temporal trends can aid in drafting targeted laws and policies that address specific risks associated with different areas and times. For instance, the identification of high-risk regions can lead to increased funding for campus safety measures and mental health resources in those areas. Moreover, understanding the demographic disparities can help in crafting policies that address the specific needs of the most affected communities, thereby fostering more inclusive and effective safety measures.

While the study provides significant insights, it is not without limitations. The generalizability of the findings may be affected by the scope of the data, which is limited to reported incidents within the United States. Furthermore, the quality of the data, especially regarding the completeness and accuracy of the incident reports, could impact the analysis. These limitations suggest that while the findings are indicative, they should be interpreted with caution and considered as part of a larger body of research.

To build on the findings of this study, future research could explore several promising avenues. Expanding the analysis to include international data could help compare the effectiveness of different campus safety policies across cultural and legal frameworks. Incorporating qualitative methods to gather more in-depth data on the circumstances surrounding campus killings could enhance understanding of the causative factors. Conducting longitudinal studies to track changes over time in response to specific policy implementations could provide insights into the effectiveness of various interventions. Exploring the role of technology in predicting and preventing campus violence could be another frontier, particularly in the age of digital surveillance and predictive analytics.

This research underscores the critical need for informed and targeted approaches to combat campus violence. By leveraging the power of data analysis, we can gain valuable insights that not only illuminate the current landscape but also pave the way for safer educational environments. The journey from data to policy is complex and requires collaborative efforts among researchers, policymakers, educators, and communities. Together, these stakeholders can harness the insights from such studies to create impactful changes that protect and enhance the lives of students across the nation and beyond.

Declarations

Author Contributions

Conceptualization: A.B.P.; Methodology: B.M.A.; Software: B.M.A.; Validation: A.A.; Formal Analysis: A.B.P.; Investigation: B.M.A.; Resources: A.A.; Data Curation: A.B.P.; Writing Original Draft Preparation: B.M.A.; Writing Review and Editing: A.B.P.; Visualization: A.B.P.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the corresponding author.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] B. Beaujolais, C. Mengo, and S. Karandikar, “An Exploration of Student, Staff, and Faculty Perceptions on the Nature of Violence and Its Prevention at a University in Turkey,” *Int. Soc. Work*, vol. 64, no. 4, pp. 496–510, 2019, doi: 10.1177/0020872819835588.

[2] R. I. Suryani, “Preventing Sexual Violence Against Women Through the Short Film ‘Demi Nama Baik?,’” *Ultim. J. Komun. Vis.*, vol. 16, no. 2, pp. 226–233, 2023, doi: 10.31937/ultimart.v16i2.3436.

[3] S. Gukurume, “Policing Toxic Masculinities and Dealing With Sexual Violence on Zimbabwean University Campuses,” *HTS Teol. Stud. Theol. Stud.*, vol. 79, no. 3, 2023, doi: 10.4102/hts.v79i3.8625.

[4] A. M. Mournet, “Associations Between Feelings of Safety, Concerns About Firearms on Campus and Suicidal Thoughts and Behaviours Among College Students,” *Inj. Prev.*, vol. 30, no. 2, pp. 171–175, 2023, doi: 10.1136/ip-2023-045060.

[5] M. ElSherief *et al.*, “Impacts of School Shooter Drills on the Psychological Well-Being of American K-12 School Communities: A Social Media Study,” *Humanit. Soc. Sci. Commun.*, vol. 8, no. 1, 2021, doi: 10.1057/s41599-021-00993-6.

[6] R. Fielding-Miller, F. Shabalala, S. Masuku, and A. Raj, “Epidemiology of Campus Sexual Assault Among University Women in Eswatini,” *J. Interpers. Violence*, vol. 36, no. 21–22, pp. NP11238–NP11263, 2019, doi: 10.1177/0886260519888208.

[7] A. Riwanto, “Addressing Campus Sexual Violence: A Collaborative Governance Approach to Legal Policy,” *Volksgeist J. Ilmu Huk. Dan Konstitusi*, pp. 225–244, 2023, doi: 10.24090/volksgeist.v6i2.9523.

[8] M. Faidah, “Strengthening the Ecosystem for Preventing and Handling Sexual Violence on Campus to Support Achievements SDGs Gender Equality,” *E3s Web Conf.*, vol. 450, p. 01007, 2023, doi: 10.1051/e3sconf/202345001007.

[9] Y. Liang, T. Liu, T. Han, H. Ferdinando, T. Seppänen, and E. Alasaarela, “Campus Violence Detection Based on Artificial Intelligent Interpretation of Surveillance Video Sequences,” *Remote Sens.*, vol. 13, no. 4, p. 628, 2021, doi: 10.3390/rs13040628.

[10] B. A. Hollister, M. J. Scalora, S. Hoff, and A. Marquez, “Exposure to Preincident Behavior and Reporting in College Students.,” *J. Threat Assess. Manag.*, vol. 1, no. 2, pp. 129–143, 2014, doi: 10.1037/tam0000015.

[11] B. E. Hayes, E. N. O’Neal, K. A. Meeker, S. A. Steele, P. Q. Brady, and M. A. Bills, “Assessing Online Strategies Aimed at Enhancing Campus Safety,” *J. Aggress. Confl. Peace Res.*, vol. 10, no. 2, pp. 112–122, 2017, doi: 10.1108/jacpr-05-2017-0293.

[12] K. M. Edwards, M. J. Mattingly, K. J. Dixon, and V. L. Banyard, “Community Matters: Intimate Partner Violence Among Rural Young Adults,” *Am. J. Community Psychol.*, vol. 53, no. 1–2, pp. 198–207, 2014, doi: 10.1007/s10464-014-9633-7.

[13] “Intelligent Campus Informatization Management Model Based on BD and AI,” *Front. Soc. Sci. Technol.*, vol. 5, no. 13, 2023, doi: 10.25236/fsst.2023.051312.

[14] R. L. Stotzer and D. MacCartney, “The Role of Institutional Factors on on-Campus Reported Rape Prevalence,” *J. Interpers. Violence*, vol. 31, no. 16, pp. 2687–2707, 2016, doi: 10.1177/0886260515580367.

[15] J. Cusano and S. McMahon, “The Impact of Campus-Level Factors on Peers’ Perceived Ability to Support a Survivor,” *J. Interpers. Violence*, vol. 36, no. 19–20, pp. NP11106–NP11130, 2019, doi: 10.1177/0886260519880169.

[16] A. L. Coker *et al.*, “Multi - College Bystander Intervention Evaluation for Violence Prevention,” *Am. J. Prev. Med.*, vol. 50, no. 3, pp. 295–302, 2016, doi: 10.1016/j.amepre.2015.08.034.

[17] C. N. Plante, D. M. Sweet, and C. Groves, “‘A Silver Lining’: Violent Media, Perceptions of a Dangerous World, and Campus Safety Intervention.,” *J. Threat Assess. Manag.*, vol. 5, no. 2, pp. 75–83, 2018, doi: 10.1037/tam0000095.

[18] T. Si, “Problem Analysis and Application Research in Legal Judgment Prediction Based on Data Mining,” *Jes*, vol. 20, no. 4s, pp. 306–316, 2024, doi: 10.52783/jes.1916.
[19] V. L. Banyard, Y. Bencosme, A. C. Cares, and M. M. Moynihan, “How Community and Peer Perceptions Promote College Students’ Pro-Social Bystander Actions to Prevent Sexual Violence,” *J. Interpers. Violence*, vol. 36, no. 7–8, pp. 3855–3879, 2018, doi: 10.1177/0886260518777557.
[20] S. Milovanović, Z. Bogdanović, A. Labus, M. Despotović-Zrakić, and S. Mitrović, “Social Recruiting: An Application of Social Network Analysis for Preselection of Candidates,” *Data Technol. Appl.*, vol. 56, no. 4, pp. 536–557, 2022, doi: 10.1108/dta-01-2021-0021.
[21] W. Shiau, Y. K. Dwivedi, and H.-H. Lai, “Examining the Core Knowledge on Facebook,” *Int. J. Inf. Manag.*, vol. 43, pp. 52–63, 2018, doi: 10.1016/j.ijinfomgt.2018.06.006.
[22] A. Handayani, N. Satyahadewi, and H. Perdana, “Application of Classification Data Mining Technique for Pattern Analysis of Student Graduation Data With Emerging Pattern Method,” *Pattimura Int. J. Math. Pijmath*, vol. 2, no. 1, pp. 01–06, 2023, doi: 10.30598/pijmathvol2iss1pp01-06.
[23] M. I. Tabash, “Synthesizing Social Insurance Research: A Bibliometric Analysis,” *Insur. Mark. Co.*, vol. 14, no. 1, pp. 59–71, 2023, doi: 10.21511/ins.14(1).2023.06.
[24] W. Zhu, S. Yabo, S. He, and W. Pan, “Corporate Social Responsibility of Internet Enterprises Based on Data Mining,” *Int. J. Electr. Eng. Educ.*, vol. 60, no. 2_suppl, pp. 236–250, 2020, doi: 10.1177/0020720920940582.
[25] S. Rajeswari, R. Poonkodi, and C. K. CharliePaul, “Protection of Direct and Indirect Discrimination Using Prevention Methods,” *Iosr J. Comput. Eng.*, vol. 16, no. 1, pp. 60–65, 2014, doi: 10.9790/0661-16116065.
[26] S. K. Garg, “Understanding the Legal and Ethical Challenges of Cryptocurrency Mining in India: An Empirical Investigation,” *Int. J. Res. Finance Manag.*, vol. 5, no. 2, pp. 343–348, 2022, doi: 10.33545/26175754.2022.v5.i2d.305.
[27] A. Gorave and V. Kulkarni, “Discrimination Aware Data Mining in Internet of Things (IoT),” *Int. J. Comput. Appl.*, vol. 159, no. 3, pp. 39–42, 2017, doi: 10.5120/ijca2017912894.
[28] W. Pedrycz, “The Benefits and Drawbacks of Data Mining Technologies,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 10, no. 1, 2019, doi: 10.1002/widm.1344.
[29] L. Carmichael, S. Stalla-Bourdillon, and S. Staab, “Data Mining and Automated Discrimination: A Mixed Legal/Technical Perspective,” *Ieee Intell. Syst.*, vol. 31, no. 6, pp. 51–55, 2016, doi: 10.1109/mis.2016.96.
[30] Yangjiu, “Campus killings.” Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/yangjiu/campus-killings