INSTITUTE FOR HOMELAND SECURITY

Sam Houston State University

WORKPLACE VIOLENCE:

ANALYZING SOCIAL NETWORKING MESSAGES TO IDENTIFY WORKPLACE HARASSMENT CASES VIA NATURAL LANGUAGE PROCESSING

Institute for Homeland Security

Sam Houston State University

Cihan Varol

Narasimha Shashidhar

Workplace Violence: Analyzing Social Networking Messages to Identify Workplace Harassment Cases via Natural Language Processing

Cihan Varol and Narasimha Shashidhar

Department of Computer Science

Sam Houston State University

Huntsville, TX, USA

Email: cvarol@shsu.edu

Abstract

The document outlines research aimed at developing a natural language processing (NLP) framework for identifying workplace harassment within social networking messages, addressing the limitations of manual reporting systems. The project seeks to create a harassment analysis and reporting framework that can proactively identify and categorize instances of workplace harassment, fostering a safer work environment.

Key challenges include the need for advanced NLP algorithms, forensic readiness, scalability across various platforms, compliance with data privacy regulations, ethical considerations, and integration with existing organizational systems. The research also examines related works, including studies on using NLP to detect suicide risk on social media, analyze bullying discourse on Twitter, and the role of emotions and ethics in group dynamics.

The proposed solution involves corporate procedures for handling harassment cases on social media platforms, including policy communication, training, an online reporting system, and a comprehensive investigation process. The NLP mechanism for analyzing social networking messages includes data preprocessing, TFIDF analysis, and classification algorithms such as Random Forest and Linear SVC. The document also details guidelines for workplace installation, emphasizing pre-installation preparation, installation steps, post-installation maintenance, documentation, compliance, scalability, disaster recovery, and testing.

In conclusion, the project aims to enhance organizations' capabilities to detect and mitigate workplace harassment effectively by leveraging NLP techniques and ensuring forensic readiness, ethical considerations, and compliance with regulations. The future work may focus on continuous improvement of the NLP model, scalability, and adapting to new communication channels.

Keywords: Harassment, Natural Language Processing, Social Media, Workplace Violence

1. Introduction and Overview

In the modern workplace, the prevalence of harassment, discrimination, and toxic behavior has become a significant concern, leading to negative impacts on employee well-being, organizational culture, and productivity. Traditional methods of identifying workplace harassment cases often rely on manual reporting systems, which are prone to skipped-being-reported due to fear of retaliation or lack of awareness. Therefore, there is a dire need for a system that can analyze social networking messages within workplace communication platforms to detect patterns indicative of harassment or hostile behavior. Overall, this project aims to develop a framework based on a natural language processing (NLP) solution capable of efficiently identifying and categorizing instances of workplace harassment, which will help organizations to proactively address such issues and foster a safer and more inclusive work environment. To this end, we propose a forensically ready harassment reporting framework that can be utilized by companies.

2. Gap Assessment / Problem statement

Despite the growing recognition of workplace harassment as a critical issue, many organizations still lack robust systems to effectively identify and address such cases. Current methods of detecting harassment often rely on manual reporting, which can be unreliable due to underreporting and subjective interpretation. Furthermore, the rise of workplace communication via social networking platforms presents a challenge, as traditional systems may not adequately capture or analyze this data.

There is a significant gap in the availability of forensic-ready systems capable of analyzing social networking messages to identify workplace harassment cases using natural language processing (NLP) techniques. Such systems would enable organizations to proactively identify and address instances of harassment, thereby fostering a safer and more inclusive work environment.

Key challenges in this domain include the need for this study are:

- Advanced NLP algorithms tailored to detect subtle nuances indicative of harassment within social networking messages.
- Forensic readiness to ensure the collected evidence is admissible in legal proceedings, maintaining integrity and reliability.
- Scalability and compatibility with various social networking platforms and communication channels commonly used in the workplace.
- Compliance with data privacy regulations and ethical considerations to safeguard employee confidentiality and rights.
- Integration with existing organizational systems and processes to streamline reporting, investigation, and resolution of harassment cases.

Overall, this work creates a forensic-ready system for analyzing social networking messages, which will help the organizations to enhance their capacity to detect and mitigate workplace harassment effectively.

There are a number of related research conducted in this area. Similar works and their details are shared in the rest of this section.

The research article "Natural Language Processing of Social Media as Screening for Suicide Risk" discusses the feasibility of using social media data to detect individuals at risk for suicide. The study aims to demonstrate the potential of natural language processing and machine learning techniques to identify quantifiable signals around suicide attempts, allowing for the development of an automated system for estimating suicide risk. The article emphasizes the importance of social media and the potential for preventive techniques by using this technology to screen for suicide risk, identifying individuals at risk prior to their engagement with the healthcare system. The authors highlight the difficulties associated with assessing an individual's risk for suicidal behavior and the limitations of existing methods, such as the latency between the risk for suicide and the suicide attempt itself, and the reliance on individuals to disclose their wish to harm themselves to a health professional. The study also discusses ethical considerations, privacy implications, and the trade-off between privacy and prevention in implementing such technology, emphasizing the need for a thoughtful and transparent public discourse to address the cultural implications of implementation. Additionally, it explores the potential sources of bias, limitations, and ethical implications associated with the use of automated screening protocols based on social media data [1].

The study titled "Bullying Discourse on Twitter" examines bullying-related tweets using supervised machine learning to understand the sharing and disclosure of bullying experiences. The research analyzes the role of the author in bullying-related tweets, identifies different types of bullying, analyzes reasons for sharing a bullying episode on Twitter, and examines temporal patterns of bullying-related tweets. A total of 847,548 tweets collected between August 7, 2019, and March 31, 2020, were analyzed, revealing that most tweets were shared from the perspective of the victim, included both general and online bullying, and the most common reason for posting was to report or to self-disclose. The study also highlighted the impact of high-profile incidents on bullying-related tweets, showcasing how these events led to spikes in the number of tweets [2].

Imran et al. discusses the use of text summarization in the context of crisis events, focusing on the challenges and methods for generating updates relating to unfolding crisis events. The author highlights the need for incremental and temporal text summarization, as well as the development of systems optimized for standardized metrics. The work also addresses the challenges in the development of summarization algorithms for crisis events, such as the relative importance of different features and the scalability of algorithms. Additionally, it touches on domain-specific approaches and the use of information extraction from social media to transform informal language into machine-readable records. The overall goal of this work is to extract time-critical information from social media that is useful for emergency responders and affected communities in disaster situations [3].

Dinakar et al. explores the use of common-sense reasoning and affective knowledge to detect implicit insults, in online interactions. By blending these techniques and leveraging Analogy Space inference, the study aims to enhance the detection of subtle forms of cyberbullying. The research

also focuses on developing reflective user interfaces to prompt users to reflect on their behavior and choices, ultimately aiming to mitigate cyberbullying through a combination of statistical machine learning and common-sense reasoning approaches [4].

Gloor et al. delves into the intricate relationship between emotions, ethics, and performance within group dynamics. It highlights the impact of fundamental emotions like fear, anger, joy, and sadness on individual and group behavior. The research emphasizes the significance of ethical values in shaping both individual and company success, showcasing how personality traits influence emotional responses. Through the innovative Tribefinder algorithm, emotions and personality attributes are classified, shedding light on the ethical, moral, and non-ethical "tribes" based on behavior. Ultimately, the study underscores the critical role of ethics and morality in fostering a conducive environment for business prosperity and effective group collaboration [5].

“Speak Up, Fight Back! Detection of Social Media Disclosures of Sexual Harassment” paper presents research on a sophisticated language model for detecting disclosures of sexual harassment on social media, particularly Twitter, in the context of the #MeToo movement. The authors argue that generic sentence classification models lack the specificity needed to handle the subtleties of language in such disclosures. They propose the Disclosure Language Model (DLM), a three-part architecture based on ULMFiT, which includes a Language Model, a Medium-Specific (Twitter) model, and a Task-Specific classifier. The researchers created a manually annotated dataset to test their model, demonstrating its superior performance over generic deep learning models and handcrafted feature-based models. They conducted an extensive comparison with state-of-the-art models and provided a detailed error analysis to support their methodology. The study highlights the potential of using social media data to understand and address sexual harassment. The authors suggest that analyzing user reactions to disclosures could inform better campaigns for social change and provide insights into the consequences of sexual abuse. The paper also addresses ethical considerations, such as privacy, bias, and the interpretation of data from vulnerable populations. The authors acknowledge the limitations of their research and propose future work, including developing a medium-agnostic model, exploring the applicability of their system for prevention, and leveraging social network analysis for a better understanding of community interactions. The DLM's architecture is detailed, including preprocessing steps, the use of dropout techniques to prevent overfitting, and a description of the training process. The model outperforms various baselines and demonstrates the effectiveness of using medium-specific language models for complex text classification tasks. The authors make their dataset publicly available to facilitate further research on this important issue [6].

The "Identification of key cyberbullies: A text mining and social network analysis approach" paper was authored by Yoon-Jin Choi, Byeong-Jin Jeon, and Hee-Woong Kim from the Graduate School of Information, Yonsei University, Korea. The study addresses the growing issue of cyberbullying and its significant societal impact, focusing on identifying key cyberbullies by applying text mining and social network analysis (SNA) methods. In this work, the researchers collected over 650,000 posts and comments from the Korean online community Daum Agora using web crawling. They calculated the Losada ratio, a measure of positive-to-negative comments, and proposed a

cyberbullying index based on the frequency of insulting words in comments. SNA was used to analyze the relationships among users and their influence on the community. The study validated the proposed method through a real-world application and found that combining the Losada ratio, cyberbullying index, and SNA centrality indices could effectively identify key cyberbullies. These individuals were characterized by a high frequency of insulting comments and significant influence within the online community [7].

This research contributes to the literature on cyberbullying by proposing and validating a method for identifying key cyberbullies, which could help manage online communities and reduce cyberbullying. The study also compiles a dictionary of Korean insulting words for detecting malicious comments and suggests that predictive policing could use the method to proactively screen cyberbullies. The authors acknowledge limitations such as the representativeness of the data and the subjective nature of malicious comments, which may not be captured by dictionary-based text mining. They suggest future research should consider deep learning methods for text classification and detecting various IPs of individual users to improve the identification of cyberbullies [7].

3. Topic Discussion

A- Corporate Procedures Overview: Corporate procedures are in place to handle harassment cases that took place in social media platforms, ensuring proper protection, analysis, and reporting as shown in Figure below. These procedures encompass the following steps:

- **Policy Communication and Training:** Develop and disseminate explicit anti-harassment policies specifically addressing social media use. These policies should define what constitutes harassment, the consequences for such behavior, and the steps for reporting incidents. Conduct regular training sessions to educate employees on recognizing and reporting social media harassment. Training should cover how to identify harassment, the importance of reporting incidents, and the support available to victims.
- **Employee:** The process begins with an employee who experiences or witnesses harassment. Employees should be encouraged to report incidents promptly and assured of confidentiality and protection against retaliation.
- **Online Harassment Reporting System:** Implement an online harassment reporting system that allows employees to report incidents easily and confidentially. The system should support anonymous reporting to protect the identity of the complainant. Ensure the reporting system is user-friendly and accessible, providing clear instructions on how to submit a report and what information to include.
- **HR Department - Initial Assessment:** HR promptly acknowledges receipt of the report to reassure the complainant that the issue is being taken seriously. Conduct an initial review to determine the severity of the complaint and whether immediate protective actions are necessary. This step ensures that urgent cases are addressed swiftly.
- **Harassment Investigation Committee:** Form an impartial investigation committee comprising HR personnel, legal advisors, and possibly external experts. The committee should be trained to handle harassment cases sensitively and impartially. Gather evidence through social media posts, screenshots, and digital communications. Collecting comprehensive evidence is crucial

for an accurate assessment. Conduct interviews with the complainant, the accused, and any witnesses. These interviews should be conducted in a non-threatening manner to ensure accurate and complete information is obtained.

- **Analysis and Findings:** Use tools to analyze the language and context of social media posts (Section 3.C). This involves evaluating the content for signs of harassment, including abusive language, threats, and patterns of behavior. Prepare a detailed report summarizing the evidence, the investigation process, and the conclusions. This report should be kept confidential and shared only with relevant parties.
- **Resolution and Action Plan:** Based on the findings, determine appropriate disciplinary measures for the accused, which may include warnings, suspension, or termination. Communicate the outcome and actions taken to the complainant, ensuring they understand the steps that have been taken and feel supported.
- **Follow-up and Support:** Monitor the workplace and social media platforms to ensure no further harassment occurs and that the complainant does not face retaliation. Offer counseling and support services to the complainant to help them recover from the incident and feel safe in their work environment.
- **Documentation and Reporting:** Securely store all documentation related to the harassment case, including the complaint, investigation records, and resolution. Regularly report cases and trends to senior management and regulatory bodies. This helps in identifying patterns and improving policies and procedures to prevent future incidents.

B- The following forensic readiness mechanism needs to be implemented in place.

- **Forensic Readiness Policy**: Establish a forensic readiness policy outlining the steps for collecting, preserving, and analyzing digital evidence related to harassment cases.
- **Legal Framework:** Ensure all forensic activities comply with relevant laws and regulations regarding privacy, data protection, and digital evidence.
- **Employee Consent:** Obtain informed consent from employees regarding monitoring and the potential collection of their social media interactions for forensic purposes.
- **Monitoring Tools**: Deploy tools to monitor social media platforms for signs of harassment, such as flagged messages, keywords, and user reports.
- **Reporting Mechanism**: Create a secure and confidential reporting mechanism for employees to report harassment incidents, including anonymous options.
- **Digital Evidence Collection**:
  - **Identify Relevant Data**: Determine the scope of data to be collected, including messages, timestamps, user profiles, and any associated media.
  - **Automated Collection Tools**: Utilize tools and software to automate the collection of relevant data from social media platforms.
  - **Manual Collection**: Establish procedures for manual collection when automated tools are insufficient or inapplicable.
- **Preservation of Evidence**:
  - **Chain of Custody**: Implement a robust chain of custody protocol to document every step in the evidence handling process.
  - **Data Integrity**: Use hashing techniques to ensure the integrity of collected data, ensuring it remains unaltered during analysis.
- **Secure Storage**: Store collected evidence in a secure, access-controlled environment to prevent unauthorized access and tampering.
- **Documentation**: Document all findings, including a detailed timeline of events and a summary of the analysis process.

C- The following mechanism needs to be used for Analyzing Social Networking Messages to Identify Workplace Harassment via NLP

C.1. Data Preprocessing: Cleaning and Preparing the Dataset for Analysis

1. **Data Collection:**
   - **Source Identification:** Identify social networking platforms where workplace interactions occur (e.g., Slack, LinkedIn, internal social networks).
   - **Data Gathering:** Collect messages from these platforms, ensuring to include both messages flagged by users or HR as potentially harassing and normal messages for comparison.
   - **Legal and Ethical Considerations:** Ensure data collection complies with privacy laws and company policies. Obtain necessary consents and anonymize data to protect privacy.
2. **Data Cleaning:**
   - **Remove Noise:** Filter out irrelevant content such as advertisements, non-textual data, URLs, and unrelated links.
   - **Normalization:** Convert all text to lowercase to maintain consistency.
   - **Tokenization:** Break down the text into individual tokens (words, phrases).

- **Stop Word Removal:** Remove common stop words (e.g., "and", "the", "is") that do not contribute to identifying harassment.
- **Lemmatization/Stemming:** Reduce words to their root forms (e.g., "running" to "run", "better" to "good") to standardize variations.

3. **Annotation:**
   - **Manual Annotation:** Have trained annotators label the dataset, indicating whether each message is harassing or non-harassing.
   - **Automated Annotation Tools:** Use semi-supervised methods to assist in labeling, ensuring accuracy through periodic manual reviews.

C.2. TFIDF Analysis: Evaluating the Importance of Words in the Context of Harassment Cases

1. **Understanding TFIDF:**
   - **Term Frequency (TF):** Measure how often a word appears in a document. Higher frequency indicates greater importance within that document.
   - **Inverse Document Frequency (IDF):** Measure the importance of a word across the entire dataset. Rare words in many documents have higher IDF scores.
2. **TFIDF Transformation:**
   - **Compute TFIDF Scores:** Convert the cleaned and tokenized text into a TFIDF matrix. Each word's score reflects its relevance for identifying harassment.
   - **Matrix Construction:** Create a matrix where rows represent documents (messages) and columns represent terms, with TFIDF scores as values.
3. **Feature Selection:**
   - **Dimensionality Reduction:** Select features (words) with the highest TFIDF scores indicative of harassment. This reduces the complexity of the dataset while retaining meaningful information.
   - **Visualization:** Use tools like word clouds or bar charts to visualize key terms identified through TFIDF analysis.

C.3. Classification Algorithms: Utilizing Algorithms to Identify Harassment

1. **Model Selection:**
   - **Algorithm Choice:** Select classification algorithms known for their robustness in text analysis:
     - **Random Forest Classifier:** An ensemble learning method using multiple decision trees to improve classification accuracy and reduce overfitting.
     - **Linear Support Vector Classification (Linear SVC):** Effective for high-dimensional data, maximizing the margin between classes for better classification.
2. **Training the Models:**
   - **Dataset Splitting:** Divide the dataset into training and testing sets (e.g., 80% for training, 20% for testing) to ensure the model's generalizability.
   - **Training Process:** Feed the TFIDF matrix into the Random Forest and Linear SVC models to learn patterns associated with harassment.
3. **Model Evaluation:**
   - **Performance Metrics:** Evaluate model performance using accuracy, precision, recall, and F1-score to measure the effectiveness of harassment detection.

- **Confusion Matrix:** Analyze the confusion matrix to understand the true positives, true negatives, false positives, and false negatives.

4. **Hyperparameter Tuning:**
   - **Optimization Techniques:** Use Grid Search, Random Search, or Bayesian Optimization to find the best hyperparameters that improve model performance.
   - **Cross-Validation:** Implement k-fold cross-validation to ensure the model's reliability and robustness.
5. **Model Deployment:**
   - **Integration:** Deploy the best-performing model into the organization's system for real-time monitoring and analysis of social networking messages.
   - **Continuous Learning:** Regularly update and retrain the model with new data to maintain high accuracy and adapt to emerging patterns.
6. **Post-Deployment Monitoring:**
   - **Feedback Loop:** Implement a feedback mechanism where users can flag false positives and false negatives to continually improve the model.
   - **Performance Tracking:** Continuously monitor the model's performance and make adjustments as necessary to ensure it remains effective.

D- Guideline for Workplace Installation: A guideline is created to prepare the system for installation on workplace computers, ensuring smooth integration and usage within the corporate environment.

D.1. Pre-installation Preparation:

- **Assessment of Requirements**: Understand the hardware and software requirements for the system and ensure compatibility.
- **Resource Allocation**: Allocate sufficient resources (CPU, memory, storage) to accommodate data processing and model training.
- **Security Considerations**: Ensure the system setup complies with security protocols and access controls to safeguard sensitive data.

D. 2. Installation Steps:

D.2.1. Data Preprocessing:

- **Setup Data Collection Mechanism**:
  - Configure data collection tools to fetch messages from identified social networking platforms securely.
  - Ensure compliance with privacy laws and company policies regarding data collection and storage.
- **Implement Data Cleaning Procedures**:
  - Develop scripts or utilize existing tools to clean and preprocess the collected data according to the outlined steps.
  - Test the cleaning process to ensure noise removal, normalization, tokenization, stop word removal, and lemmatization/stemming are performed accurately.
- **Establish Annotation Framework**:
  - Set up a platform or utilize annotation tools for manual and automated annotation of the dataset.

- Ensure proper training for annotators and periodic review of automated annotations to maintain accuracy.

D.2.2. TFIDF Analysis:

- **TFIDF Transformation Implementation**:
  - Integrate libraries or develop scripts to compute TFIDF scores and construct the TFIDF matrix.
  - Verify the correctness of the TFIDF computation process and matrix construction.
- **Feature Selection Integration**:
  - Incorporate methods for dimensionality reduction based on TFIDF scores to select relevant features.
  - Implement visualization tools for better understanding and interpretation of selected features.

D.2.3. Classification Algorithms:

- **Select and Integrate Classification Algorithms**:
  - Set up environments for deploying Random Forest Classifier and Linear SVC algorithms.
  - Ensure necessary libraries and dependencies are installed for model training and evaluation.
- **Training and Evaluation**:
  - Develop scripts or utilize existing frameworks to train models on the prepared dataset and evaluate performance using specified metrics.
  - Implement confusion matrix analysis to assess model behavior and performance.
- **Hyperparameter Tuning and Model Deployment**:
  - Integrate optimization techniques for hyperparameter tuning and deploy the best-performing model into the corporate system.
  - Establish mechanisms for model versioning and deployment tracking.

D.3. Post-installation Maintenance and Monitoring:

- **Feedback Mechanism Implementation**:
  - Set up channels for users to provide feedback on model predictions, particularly false positives and false negatives.
  - Develop processes for incorporating user feedback into model updates.
- **Continuous Monitoring and Performance Tracking**:
  - Implement monitoring tools to track model performance metrics and system health.
  - Establish protocols for regular model retraining and updates based on new data and emerging patterns.

D.4. Documentation and Training:

- **Create User Manuals and Documentation**:
  - Prepare comprehensive documentation covering system installation steps, usage guidelines, and troubleshooting procedures.

- Conduct training sessions for system administrators and end-users to ensure effective utilization of the implemented mechanism.

D.5. Compliance and Governance:

- **Regular Audits and Compliance Checks**:
  - Schedule periodic audits to ensure ongoing compliance with privacy regulations and company policies.
  - Establish governance frameworks for responsible AI usage and ethical data handling practices.

D.6. Scalability and Adaptability:

- **Plan for Scalability**:
  - Design the system architecture with scalability in mind to accommodate future growth in data volume and user demand.
  - Evaluate potential integration with cloud-based services for scalability and resource optimization.

D.7. Disaster Recovery and Backup:

- **Implement Backup and Recovery Strategies**:
  - Set up regular backups of critical system components and data to prevent data loss in case of system failures or security breaches.
  - Develop contingency plans and procedures for restoring system functionality in the event of disruptions.

D.8. Testing and Quality Assurance:

- **Conduct Comprehensive Testing**:
  - Perform thorough testing of the installed system across different scenarios to identify and address potential issues.
  - Engage in continuous quality assurance practices to ensure the reliability and performance of the deployed mechanism.

4. Practical Applicability

The practical applicability of the proposed NLP framework for identifying workplace harassment within social networking messages is significant for several reasons: A) The system can proactively identify patterns of harassment that may not be immediately apparent to human observers, thus preventing potential escalation of harmful behavior. B) By automating the detection process, the system can address the issue of underreporting due to fear of retaliation or lack of awareness among employees. C) The NLP framework can process large volumes of data quickly, making the identification of harassment more efficient compared to manual methods. D) The use of NLP algorithms reduces the subjectivity involved in interpreting messages, providing a more objective assessment of whether harassment has occurred. E)

The forensically ready aspect of the system ensures that any evidence collected can be used in legal proceedings, maintaining the integrity and reliability of the data. F) The framework is designed to be scalable and compatible with various social networking platforms, making it adaptable to different organizational structures and communication channels. G) The solution can be integrated with existing organizational systems, allowing for streamlined reporting, investigation, and resolution of harassment cases. H) The system is designed with compliance with data privacy regulations and ethical considerations in mind, ensuring employee confidentiality and rights are protected. I) The NLP model can be continuously updated and retrained with new data, enabling it to adapt to evolving language patterns and maintain its effectiveness over time. J) The framework includes provisions for follow-up and support services for victims, helping them recover from the incident and ensuring a safe work environment.

5. Way Forward

This work proposed an innovative NLP framework designed to proactively identify and categorize instances of workplace harassment within social networking messages. This framework addresses the shortcomings of traditional manual reporting systems by leveraging advanced NLP algorithms, ensuring forensic readiness, and complying with data privacy regulations and ethical standards. The proposed solution includes corporate procedures for handling harassment cases, an online reporting system, and a comprehensive investigation process supported by NLP mechanisms such as data preprocessing, TFIDF analysis, and classification algorithms. By integrating with existing organizational systems, the framework aims to streamline the detection and mitigation of workplace harassment, fostering a safer and more inclusive work environment.

Future work on the NLP framework for identifying workplace harassment could focus on several areas. First, continuous improvement of the NLP model is essential to adapt to evolving language patterns and maintain high accuracy in harassment detection. This could involve incorporating feedback loops to refine the model based on user input and false positive/negative rates. Secondly, scalability is a key aspect that requires further development to ensure the framework can handle increasing volumes of data and accommodate growth in user base and communication channels. Additionally, adapting the framework to new and emerging social media platforms and communication channels will be necessary to stay relevant and effective. Ethical considerations and privacy concerns must remain at the forefront, with ongoing research into best practices for balancing the need for surveillance with respect for individual privacy. Finally, the framework could be expanded to include additional features such as sentiment analysis, context-aware processing, and multilingual capabilities to enhance its applicability across diverse organizational settings and global workforces.

6. References

[1] Coppersmith, G., Leary, R., Crutchley, P. Fine, A. Natural language processing of social media as screening for suicide risk. Biomed. Inform. Insights, 10 (2018), Article 117822261879286, 10.1177/1178222618792860

[2] Dhungana Sainju K., Mishra N., Kuffour A., Young L. Bullying discourse on Twitter: An examination of bully-related tweets using supervised machine learning. Comput. Hum. Behav. 2021;120:106735. doi: 10.1016/j.chb.2021.106735

[3] Imran M, Castillo C, Diaz F, Vieweg S. Processing social media messages in mass emergency: A survey. ACM Computing Surveys (CSUR). 2015;47(4):1–38. doi: 10.1145/2771588

[4] Dinakar, K., Jones, B., Havasi, C., Lieberman, H., Picard, R. Commonsense Reasoning for Detection, Prevention and Mitigation of Cyberbullying. BodyNets International Conference on Body Area Networks. vol. 1, no. 212, 2012.

[5] Gloor, P., Fronzetti Colladon, A., Grippa, F. Measuring ethical behavior with AI and natural language processing to assess business success. Scientific Reports, 12(1), 1-13, 2022

[6] Chowdhury, A., Sawhney, R., Mathur, P., Mahata, D., Shah, R.R. Speak up, Fight Back! Detection of Social Media Disclosures of Sexual Harassment. 136-146. 10.18653/v1/N19-3018, 2019

[7] Choi, Y.J., Jeon, B.J., Kim, H.W. Identification of key cyberbullies: A text mining and social network analysis approach. Telematics and Informatics. 20205620210.1016/j.tele.101504

Biography

Dr. Cihan Varol, is a Professor of Computer Science at Sam Houston State University. He received his Bachelor of Science degree in Computer Science from Firat University, Elazig, Turkey in 2002, Master of Science degree from Lane Department of Computer Science and Electrical Engineering from West Virginia University, Morgantown, WV, USA in 2005, and Doctor of Philosophy in Applied Computing from University of Arkansas at Little Rock in 2009. His research interests are in the general area of information (data) quality and its applications on Digital Forensics and Cyber Security areas, with specific emphasis on personal identity recognition, privacy preserving record linkage, entity resolution, secured IoT systems, social media forensics, 3D printer forensics, and web forensics. These studies have led to more than 130 peer-reviewed journal and conference publications and three book chapters. He is an executive board member of IEEE Education Society Standards Committee and the chair of IEEE P2834.1 Standards on Digital Forensics on Trusted Learning Systems.

Dr. Narasimha Shashidhar, Professor and Director for the Doctor of Philosophy Program in Digital and Cyber Forensic Science. Dr. Narasimha Shashidhar received his Bachelor of Engineering in Electronics and Communication Engineering from The University of Madras in 2001, and his M.S. and Ph.D. degrees in Computer Science and Engineering from The University of Connecticut in 2004 and 2010, respectively. His research interests include Digital Forensics, Information Security, Cyber Forensics, and Computing Education. He was a part of the Voting Technology and Research Center (VoTeR) at the University of Connecticut where he advised the State of CT on the security and deployment of electronic voting

machines. He has over 100 conference/journal publications and serves on the editorial advisory/review board and the Technical Program Committee (TPC) of a number of books, journals, and conferences.

INSTITUTE FOR HOMELAND SECURITY

Sam Houston State University

The Institute for Homeland Security at Sam Houston State University is focused on building strategic partnerships between public and private organizations through education and applied research ventures in the critical infrastructure sectors of Transportation, Energy, Chemical, Healthcare, and Public Health.

The Institute is a center for strategic thought with the goal of contributing to the security, resilience, and business continuity of these sectors from a Texas Homeland Security perspective. This is accomplished by facilitating collaboration activities, offering education programs, and conducting research to enhance the skills of practitioners specific to natural and human caused Homeland Security events.

Institute for Homeland Security
Sam Houston State University

Varol, Cihan & Shashidhar, Narasimha (2024) Workplace Violence: Analyzing Social Networking Messages to Identify Workplace Harassment Cases via Natural Language Processing (Report No. IHS/CR-2024-1009). The Sam Houston State University Institute for Homeland Security.
https://doi.org/10.17605/OSF.IO/YD34B