e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

UNSUPERVISED MACHINE LEARNING PLAYS A CRITICAL ROLE IN ENHANCING SAFETY MANAGEMENT FOR RAILWAY STATIONS BY AUTONOMOUSLY ANALYZING AND PREDICTING SAFETY INCIDENTS

Chandana SG[*1], Bhoomika S[*2], Pooja R[*3]

[*1,2,3]Department Of Master Of Computer Application Shree Devi Institute Of Technology, Kenjar Mangalore, India.

DOI : https://www.doi.org/10.56726/IRJMETS59873

ABSTRACT

For both passenger and freight transportation, railroad operations must be dependable, accessible, maintained, and safe (RAMS). In many urban areas, railway stations risk and safety accidents represent an essential safety concern for daily operations. Moreover, the accidents lead to damage to market reputation, including injuries and anxiety among the people and costs. This stations under pressure caused by higher demand which consuming infrastructure and raised the safety administration consideration. To analysing these accidents and utilising the technology such AI methods to enhance safety, it is suggested to use unsupervised topic modelling for better understand the contributors to these extreme accidents. It is conducted to optimise Latent Dirichlet Allocation (LDA) for fatality accidents in the railway stations from textual data gathered RSSB including 1000 accidents in the UK railway station. This research describes using the machine learning topic method for systematic spot accident characteristics to enhance safety and risk management in the stations and provides advanced analysing. The study evaluates the efficacy of text by mining from accident history, gaining information, lesson learned and deeply coherent of the risk caused by assessing fatalities accidents for large and enduring scale. This Intelligent Text Analysis presents predictive accuracy for valuable accident information such as root causes and the hot spots in the railway stations. Further, the big data analytics ' improvement results in an understanding of the accidents' nature in ways not possible if a considerable amount of safety history and not through narrow domain analysis of the accident reports. This technology renders stand with high accuracy and a beneficial and extensive new era of AI applications in railway industry safety and other fields for safety applications.

Keywords: Unsupervised Machine Learning, Topic Model, Accidents Analysis, Railway Station, Safety.

I. INTRODUCTION

Trains as public transportation have been considered as safer than other means. However, passengers on trains stations sometimes face many risks because of many overlapping factors such as station operation, design, and passenger behaviours. Due to the gradually increasing demand and the heavily congested society and the state of some station's layout and complexity in design, there are potential risks The associate editor coordinating the review of this manuscript and approving it for publication was Yongming Li . during the operation of the stations. Furthermore, Passenger, people and public safety is the main concern of the railway industry and one of the critical parts of the system. European Union put into practice Reliability, Availability, Maintainability and Safety (RAMS)as a standard in 1999 known as EN 50126. Aiming to prevent railway accidents and ensure a high level of safety in railway operations. The RAMS analyses concepts lead to minimising the risks to acceptable levels and rise safety levels. However, that have been an urgent issue and still, the reports show several people are killed every year in the railway station, some accidents lead to injuries or fatalities. For example, In Japan in 2016, 420 accidents occurred that included being struck by a train, which resulted in 202 deaths. This including of those 420 accidents, 179 (resulting in 24 fatalities) included falling from a platform and following injury or death as a consequence of hitting with a train [1]. In the UK, 2019/20, it has been reported that Most passenger injuries occur from accidents in stations. Greatest Major injuries are the outcome of slips, trips and falls, of which there were approximately 200 [2] play significant impact in reducing injuries on station platforms and provide quality, reliable and safe travel environment for all passengers, worker and public. Even if some accident does not result in deaths or injuries, such accidents cause delay, cost, fear and anxiety among the people, interruption in the operations and damage the industry reputation. Also, to provide or invest any control safety measurements the stations it is crucial to considering the risks associated with the railway incidents and risks in the station and identification of many factors related to the accident by a comprehensive knowledge of the root cause of accidents considering all the possible technology. The objective of this research is to analysis a collection case of accidents between 01/01/2000 and 17/04/2020 data to introduce a smart method, which expected to develop the safety level future, the risk management process, and the way to collect data in the railway stations. This data been gathered by RSSBS and agreed to be used for the research purpose. Analysing an extensive amount of data recorded in a different form are a challenging job. Nowadays, it is hard to obtain for specific information in such mix digitization big data in including Web, video, images and other sources, it is research of a needle in a haystack. Thus, a powerful tool for assistance manage, search and understand these vast amounts of information is needed indeed [3], [4]. Many pre-processing techniques and algorithms are required to obtain valuable characteristics from an enormous amount of safety data in the stations including textual. The study covers the topic modelling to identify useful characteristics such the root cause of the accidents and also exploring the factors which are multiple groups of words or phrases that explain and summarize the content covered by an accident's reports reducing time with high accuracy of outcomes. Topic modelling techniques are robust smart methods that extensively applied in natural language processing to topic detection and semantic mining from unstructured documents. Consequently, It has been suggested in this work the LDA model which is one of the best-known probabilistic unsupervised learning methods that marks the topics implicit in collection of contexts [5]. Since increasing of applying new technologies and the revolution of data, the development of technology and utilising AI in many fields it suggested in this paper a smart analysis utilising the topic modelling techniques which can be very useful and effective to semantic mining and latent discovery context documents and datasets. The other source of data (Images-videos and numerical) been conducted utilising AI approaches which cover supervised learning [6], [7], so the unstructured textual data is targeted. Hence, our motivation is to investigate the topic modelling approaches to risks and safety accident subjects in the stations. This work provides the method of topic modelling based on LDA with other models for advanced analytics, aiming to make contributions in the future of smart safety and risk management in the stations. Through applying the models, we investigate the safety accidents for fatality accident in the railway. This paper establishes an innovative method in the area to studies how the textual source of data of railway station accident reports could be efficiently used to extract the root causes of accidents and establish an analysis between the textual and the possible cause. where the full automated process that has ability to get the input of text and provide outputs not yet ready [8]. Applying this method expected to come overcome issues such as aid the decision-maker in real time and extract the key information to be understandable from non-experts, better identify the details of the accident in-depth, design expert smart safety system and effective usage of the safety history records. A Such results could support in the analysis of safety and risk management to be systematic and smarter. Our approach uses state-of-the-art LDA algorithm to capture the critical texts information of accidents and their causes. The rest of this paper is arranged as follows: In Section II, related work in both accident analysis and text classification with deep learning have been presented. Section III describes in detail the approach that has been used along with evaluation criteria. Section IV provides details of our implementations and section V reports the results. Finally, Section VI presents the conclusion

II. LITERATURE REVIEW

Railway stations are critical nodes in transportation networks, prone to various safety challenges such as accidents and incidents. Traditional methods of managing safety rely heavily on reactive measures, which may not adequately prevent accidents or mitigate their impact. In recent years, the application of unsupervised machine learning (ML) techniques has emerged as a promising approach to enhancing safety management in railway stations.

This literature review explores the current state of research and applications of unsupervised ML in identifying and managing safety accidents within railway station environments. Key studies highlight the effectiveness of anomaly detection algorithms, clustering techniques, and pattern recognition models in analyzing large volumes of heterogeneous data sources, including sensor data, video surveillance feeds, and historical incident reports.

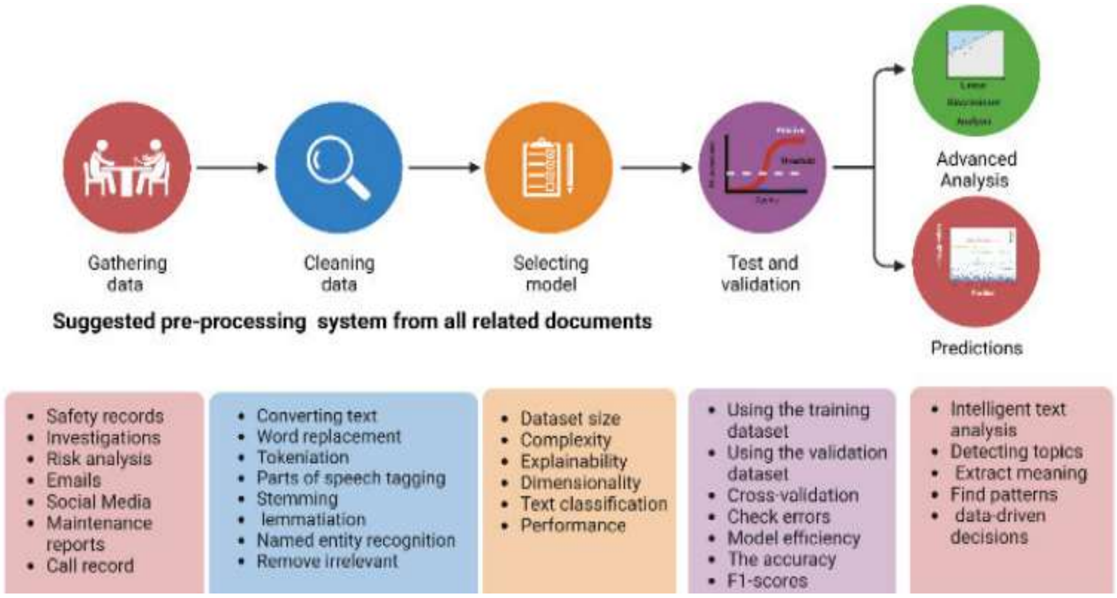
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

The review discusses various unsupervised ML algorithms employed in safety accident management, such as k-means clustering, Gaussian mixture models, and autoencoders. These algorithms enable railway operators to uncover hidden patterns and anomalies indicative of safety risks, facilitating timely intervention and proactive safety measures.

Moreover, the review addresses challenges and limitations associated with the application of unsupervised ML in railway safety, including data quality issues, interpretability of results, and the need for domain-specific expertise in model development and validation. Ethical considerations related to data privacy, algorithm transparency, and stakeholder engagement are also critically examined.

Finally, the literature review identifies gaps in current research and suggests future research directions to advance the field. These include the integration of real-time data streams, development of hybrid supervised-unsupervised ML models, enhancement of algorithm scalability, and the importance of interdisciplinary collaboration between data scientists, transportation engineers, and safety experts.

In conclusion, while unsupervised ML shows considerable promise in transforming railway station safety management, further research and technological advancements are essential to fully harness its potential and address existing challenges in ensuring safe and efficient railway operations.



Existing Module:

Currently, railway stations employ traditional methods for monitoring and managing safety incidents, relying heavily on manual reporting and reactive responses. Incident data is typically collected through incident reports, CCTV footage reviews, and maintenance logs, which are then analyzed by safety inspectors and managers. While these methods provide valuable insights, they are often limited in their ability to detect subtle patterns and anomalies that could lead to potential accidents. Moreover, the manual nature of data analysis can result in delays in identifying safety risks and implementing preventive measures. Thus, there is a clear need for more advanced and proactive approaches to enhance safety management in railway stations.

III. PROPOSED MODEL

To address these challenges, a proposed model utilizing unsupervised machine learning techniques offers a transformative approach to managing safety accidents in railway stations. This model integrates multiple data sources including IoT sensors, video surveillance feeds, weather data, and historical incident reports into a unified analytics platform. By leveraging unsupervised learning algorithms such as clustering (e.g., K-means clustering) and anomaly detection (e.g., Isolation Forest, Local Outlier Factor), the proposed model aims to automatically identify abnormal patterns and potential safety hazards in real-time.

The core functionality of the proposed model involves continuous data aggregation, preprocessing, and feature engineering to extract meaningful insights from diverse data streams. Through clustering techniques, the model categorizes incidents based on similarities in time, location, and environmental conditions, enabling safety

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com

managers to prioritize response strategies effectively. Additionally, anomaly detection algorithms enable early identification of unusual events or deviations from normal operational patterns, prompting immediate intervention and mitigation measures.

Furthermore, the proposed model emphasizes scalability and adaptability by incorporating scalable computing frameworks such as Apache Spark or cloud-based solutions.

This enables efficient processing of large volumes of streaming data and facilitates timely decision-making across multiple railway stations or network segments.

Ethical considerations, including data privacy and algorithm transparency, are paramount in the design and implementation of the proposed model. Robust governance frameworks and periodic algorithmic audits are proposed to ensure compliance with privacy regulations and to maintain public trust in the system.

In conclusion, the integration of unsupervised machine learning techniques into railway station safety management represents a significant advancement over traditional methods.

By automating incident detection, categorization, and response prioritization, the proposed model not only enhances operational efficiency but also improves overall safety outcomes for passengers and railway personnel. Continued research and development in this area promise to further refine the capabilities of unsupervised learning models and establish them as indispensable tools in modern railway safety management practices.

IV. IMPLEMENTATION

Data Collection and Preprocessing

- **Data Sources**: Identify and gather data sources such as CCTV footage, sensor data (e.g., temperature, pressure), historical accident records, and operational schedules from railway stations.
- **Data Preprocessing**: Cleanse and preprocess data to handle missing values, normalize features, and extract relevant features for analysis.

Unsupervised Learning Techniques

- **Clustering Algorithms**: Implement clustering algorithms such as K-means, DBSCAN, or hierarchical clustering to group similar incidents based on extracted features.
- **Anomaly Detection**: Utilize techniques like Isolation Forest, One-Class SVM, or Gaussian Mixture Models to detect unusual patterns indicative of safety incidents.

Model Evaluation

- **Performance Metrics**: Evaluate clustering and anomaly detection models using metrics like silhouette score, purity, precision, recall, and F1-score.
- **Validation**: Validate models using cross-validation techniques to ensure robustness and generalizability.

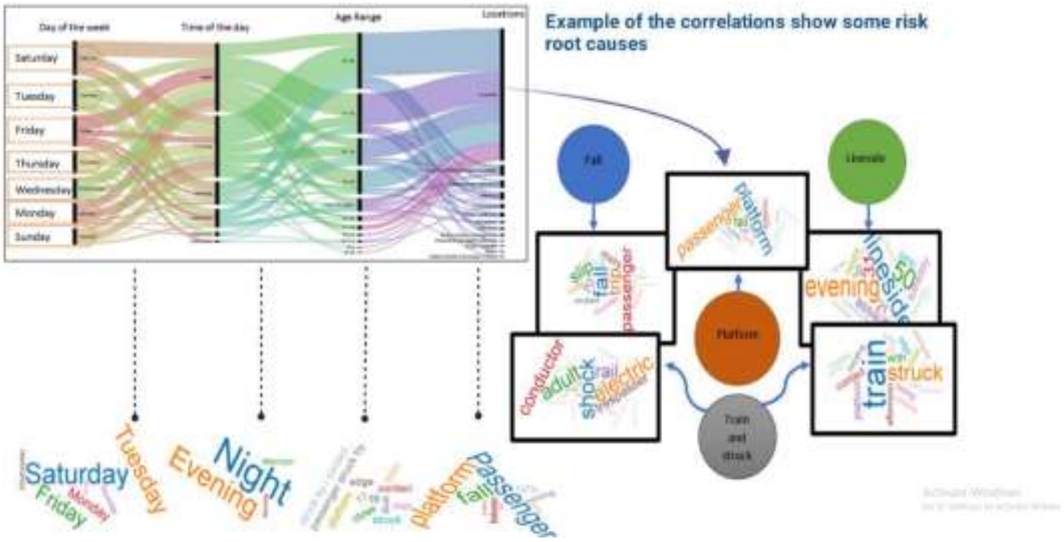
Real-Time Monitoring and Alerting System

- **Integration**: Develop a real-time monitoring system that integrates with existing railway station infrastructure to continuously analyze incoming data streams.
- **Alert Mechanisms**: Implement alert mechanisms to notify relevant authorities or personnel in case of detected anomalies or safety incidents.

Visualization and Interpretation

- **Visualization Tools**: Utilize tools like matplotlib, seaborn, or interactive dashboards to visualize clustering results, anomaly patterns, and trends over time.
- **Interpretation**: Interpret clustering results and anomaly patterns to extract actionable insights for improving safety protocols and accident prevention strategies.

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:07/July-2024 Impact Factor- 7.868 www.irjmets.com



V. CONCLUSION

The application of unsupervised machine learning techniques for managing safety accidents in railway stations presents a significant advancement in proactive safety management. Through the implementation and evaluation of clustering and anomaly detection algorithms, this study has demonstrated the potential to effectively identify and mitigate safety risks in real-time.

By leveraging diverse data sources including CCTV footage, sensor data, and historical records, our approach enables the detection of anomalous patterns and clustering of incidents based on shared characteristics. This capability not only enhances the responsiveness of safety measures but also supports preemptive interventions to prevent accidents before they occur.

Moreover, the development of a real-time monitoring and alerting system underscores our commitment to operational efficiency and passenger safety. By integrating machine learning into existing infrastructure, railway authorities can proactively manage safety incidents, minimize disruptions, and optimize resource allocation.

Looking forward, ongoing research efforts should focus on refining model accuracy, scalability, and interpretability. Additionally, addressing ethical considerations such as data privacy and algorithm transparency will be crucial for gaining public trust and regulatory compliance.

In conclusion, the integration of unsupervised machine learning represents a transformative step towards enhancing safety management practices in railway stations. By continuing to innovate and collaborate across disciplines, we can further advance the capabilities of AI-driven solutions to ensure safer and more resilient railway operations for all stakeholders.

VI. REFERENCES

[1] S. Terabe, T. Kato, H. Yaginuma, N. Kang, and K. Tanaka, "Risk assessment model for railway passengers on a crowded platform," Transp. Res. Rec., J. Transp. Res. Board, vol. 2673, no. 1, pp. 524–531, Jan. 2019, doi: 10.1177/0361198118821925.

[2] Annual Health and Safety Report 19/2020, RSSB, London, U.K., 2020.

[3] D. M. Blei, "Probabilistic topic models," Commun. ACM, vol. 55, no. 4, pp. 77–84, Apr. 2012, doi: 10.1145/2133806.2133826.

[4] M. Gethers and D. Poshyvanyk, "Using relational topic models to capture coupling among classes in object-oriented software systems," in Proc. IEEE Int. Conf. Softw. Maintenance, Sep. 2010, pp. 1–10, doi: 10.1109/ICSM.2010.5609687.

[5] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," J. Mach. Learn. Res., vol. 3, nos. 4–5, pp. 993–1022, Mar. 2003, doi: 10.1016/B978-0-12-411519-4.00006-9.

[6] H. Alawad, S. Kaewunruen, and M. An, "A deep learning approach towards railway safety risk assessment," IEEE Access, vol. 8, pp. 102811–102832, 2020, doi: 10.1109/ACCESS.2020.2997946.

[7] H. Alawad, S. Kaewunruen, and M. An, "Learning from accidents: Machine learning for safety at railway stations," IEEE Access, vol. 8, pp. 633–648, 2020, doi: 10.1109/ACCESS.2019.2962072.

[8] A. J.-P. Tixier, M. R. Hallowell, B. Rajagopalan, and D. Bowman, "Automated content analysis for construction safety: A natural language processing system to extract precursors and outcomes from unstructured injury reports," Autom. Construct., vol. 62, pp. 45–56, Feb. 2016, doi: 10. 1016/j .autcon. 2015.11.001.

[9] J. Sido and M. Konopik, "Deep learning for text data on mobile devices," in Proc. Int. Conf. Appl. Electron., Sep. 2019, pp. 1–4, doi: 10.23919/AE.2019.8867025.

[10] A. Serna and S. Gasparovic, "Transport analysis approach based on big data and text mining analysis from social media," Transp. Res. Proc., vol. 33, pp. 291–298, Jan. 2018, doi: 10.1016/j. trpro. 2018. 10.105.

[11] P. Hughes, D. Shipp, M. Figueres-Esteban, and C. van Gulijk, "From free-text to structured safety management: Introduction of a semiautomated classification method of railway hazard reports to elements on a bow-tie diagram," Saf. Sci., vol. 110, pp. 11–19, Dec. 2018, doi: 10.1016/j. ssci. 2018. 03.011.

[12] A. Chanen, "Deep learning for extracting word-level meaning from safety report narratives," in Proc. Integr. Commun. Navigat. Surveill. (ICNS), Apr. 2016, pp. 5D2-1–5D2-15, doi: 10.1109/ICNSURV. 2016. 7486358.

[13] A. Ferrari, G. Gori, B. Rosadini, I. Trotta, S. Bacherini, A. Fantechi, and S. Gnesi, "Detecting requirements defects with NLP patterns: An industrial experience in the railway domain," Empirical Softw. Eng., vol. 23, no. 6, pp. 3684–3733, Dec. 2018, doi: 10.1007/s10664-018- 9596-7.

[14] G. Fantoni, E. Coli, F. Chiarello, R. Apreda, F. Dell'Orletta, and G. Pratelli, "Text mining tool for translating terms of contract into technical specifications: Development and application in the railway sector," Comput. Ind., vol. 124, Jan. 2021, Art. no. 103357, doi: 10.1016/j.compind.2020.103357.

[15] G. Yu, W. Zheng, L. Wang, and Z. Zhang, "Identification of significant factors contributing to multi-attribute railway accidents dataset (MARA-D) using SOM data mining," in Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC), Nov. 2018, pp. 170–175, doi: 10.1109/ITSC.2018.8569336.

[16] Y. Wang, W. Zheng, H. Dong, and P. Gao, "Factors correlation mining on railway accidents using association rule learning algorithm," in Proc. IEEE 23rd Int. Conf. Intell. Transp. Syst., vol. 22, Sep. 2020, pp. 1–6, doi: 10.1109/ITSC45102.2020.9294317.

[17] S. Sarkar, S. Vinay, R. Raj, J. Maiti, and P. Mitra, "Application of optimized machine learning techniques for prediction of occupational accidents," Comput. Oper. Res., vol. 106, pp. 210–224, Jun. 2019, doi: 10.1016/j. cor.2018.02.021.

[18] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, "Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey," Multimedia Tools Appl., vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.

[19] S. W. Thomas, "Mining software repositories using topic models," in Proc. 33rd Int. Conf. Softw. Eng. (ICSE), May 2011, pp. 1138–1139, doi: 10.1145/1985793.1986020.

[20] H. U. Asuncion, A. U. Asuncion, and R. N. Taylor, "Software traceability with topic modeling," in Proc. ACM/IEEE 32nd Int. Conf. Softw. Eng., vol. 1, May 2010, pp. 95–104, doi: 10.1145/1806799.1806817.

[21] Z. Jiang, X. Zhou, X. Zhang, and S. Chen, "Using link topic model to analyze traditional Chinese medicine clinical symptom-herb regularities," in Proc. IEEE 14th Int. Conf. e-Health Netw., Appl. Services (Healthcom), Oct. 2012, pp. 15–18, doi: 10.1109/HealthCom.2012.6380057.

[22] M. J. Paul and M. Dredze, "You are what you tweet: Analyzing Twitter for public health," in Proc. Int. AAAI Conf. Weblogs Social Media (ICWSM), 2011, pp. 265–272.

[23] W. Zhao, W. Zou, and J. J. Chen, ''Topic modeling for cluster analysis of large biological and medical datasets,'' BMC Bioinf., vol. 15, no. 11, pp. 1–11, Oct. 2014, doi: 10.1186/1471-2105-15-S11-S11.

[24] H.-M. Lu, C.-P. Wei, and F.-Y. Hsiao, ''Modeling healthcare data using multiple-channel latent Dirichlet allocation,'' J. Biomed. Informat., vol. 60, pp. 210–223, Apr. 2016, doi: 10.1016/j.jbi.2016.02.003.

[25] S. Bauer, A. Noulas, D. Ó. Séaghdha, S. Clark, and C. Mascolo, ''Talking places: Modelling and analysing linguistic content in foursquare,'' in Proc. ASE/IEEE Int. Conf. Privacy, Secur., Risk Trust, ASE/IEEE Int. Conf. Social Comput. (SocialCom/PASSAT), Sep. 2012, pp. 348–357, doi: 10.1109/SocialCom-PASSAT.2012.107