Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Framework for extraction of crime insights from diversified data sources using a multiple minimum support FP-Growth Algorithm

George Matto[1], Lucas Mjema[2]
[1,2] *Department of ICT, Moshi Co-operative university, Tanzania*
[1] *Corresponding author's Email: george.matto@mocu.ac.tz*

ARTICLE INFO	ABSTRACT

Received: 30 Dec 2024

Revised: 12 Feb 2025

Accepted: 26 Feb 2025

Introduction: Mining of crime insights from datasets can be employed to help in discovering useful crime insights for improved crime detection and prevention strategies. However, most of existing frequent pattern mining approaches lack flexibility in terms of user-defined minimum supports and diversified sources of input data.

Objectives: This study was carried out to propose a framework for extaction of crime insights from diversified sources of data using the FP-Growth algorithm with multiple minimum supports.

Methods: The study was based on systematic review of literature to establish the study gap followed by a comprehensive experimentation for the validation of the proposed framework. With regard to the systematic review, a PRISMA framework was employed. To accomplish experimentation of the suggested framework, two different sources of data were used; crimes database which provided structured data, and news articles which provided unstructured data.

Results: The study came up with a generic and flexible framework that extracts crime insights from diversified data sources composing of four stages as follows: data sources, pre-processing, processing, and pattern visualization.. On top of being effective in the extraction of patterns, this approach yields a better runtime and memory use than classical FP-growth

Conclusions: Multiple sources of crime data should be considered for effective extraction of crime insights. For it to be effective, frequent pattern mining approach must consider using multiple minimum supports.

Keywords: Crime, Crime insights, Framework, FP-Growth, Multiple minimum support

INTRODUCTION

The effects of crime occur in both developed and developing countries, forcing countries to raise expenditure on security (Sugiharti et al., 2022). Matto (2019) indicated that such effects are on both crime victims and the public. Crime victims suffer bodily injury or financial loss as an instant consequences, and emotional and psychological problems as long-term effects. With regard to the community and public at large, crime reduces the perception of safety, disrupts order, creates chaos, and causes economic costs. Across the globe, countries are endeavoring to ensure crime is prevented and detected (Vo et al., 2020; De Farias et al., 2018). Unfortunately, mechanisms to detect and prevent crime in many less-developed countries are based on inadequate solutions and practices. Mussa (2023) mentioned reliance on classical closed-circuit television cameras, electric fencing, high walls, private security guards, security dogs, police patrol, youth engagement, and community policing as examples of crime prevention practices in many developing countries.

As the world has entered into the fifth industrial revolution, which is characterized by advanced technologies and artificial intelligence, criminals plan, organize, and in some cases execute their actions with the help of technologies. In that case, traditional approaches and practices fail to prevent, detect or capture the criminals. But as said by Matto (2019), most of criminal activities that are done via technological platforms leave traces through data they generate.

For instance, they leave traces through phone communications, emails, social media, click streams, web logs, IP addresses and location data. Likewise, due to increased use of technology, most members of the public report crimes that occur in their communities with the aid of technologies such as phone calls, SMS, crime reporting systems as well as social media and other online platforms (Oghogho et al., 2024; Maslennikova et al., 2021). In that way, they generate data that once that analyzed can provide helpful crimes insights. Further, now-a-days, police departments maintain crime records in electronic platforms (Chowdhury, 2023; Dekker et al., 2020). Such records can also be analyzed to provide useful crime insights necessary for improving crime detection and prevention strategies.

There are numerous approaches for extracting frequent patterns from datasets (Fournier-Viger et al., 2022; Nasyuha et al., 2020; Min et al., 2020). A frequent pattern is a pattern in the dataset that occurs with a frequency not less than a certain specified threshold (Mokkadem at al., 2023). For example, if three is the minimum threshold, all items appearing together in the dataset three times or more are regarded are frequent itemsets. Discovering frequent itemsets paves a way to the extraction of association rules and several other compelling relationships between items in the dataset. An association rule is an implication of the form A→B, where A is the antecedent, and B is the consequent (Matto and Mwangoka, 2018). This gives information on the form; when A occurs, B is likely to occur as well, which is very important in uncovering hidden patterns in large datasets (Mokkadem et al., 2023). For instance, an association rule might reveal that "75% of bank robbers, between 25 and 45 years old, kill their victims". This could mean that youth criminals involved in robbery in the bank end up killing the victims. Thus, the mined frequent patterns help to uncover vital information regarding criminals' behavior and actions.

Several frequent itemset mining and association rules generation approaches have been not only proposed but also tried-out. The Apriori (Agrawal and Srikant, 1994), Eclat (Zaki et al., 1997), and FP-Growth algorithm (Han et al., 2004) are some of the most widely used algorithms for association rules mining. Research has established FP-Growth to be more effective as compared to other association-rule-mining algorithms. This algorithm has also been widely used for crime pattern mining (Macingwane & IsafIade, 2023; Franchina et al., 2022; Roy et al., 2021). However, many studies and experiments on FP-Growth have focused on classical FP-Growth, which extracts frequent patterns based on a single minimum support defined by the user. This technique is not suitable for crime-pattern mining because of the complexion of items in the crime dataset in which some crime items appear frequently because they are frequently committed while others seldomly emerge.

Likewise, several frameworks for crime-pattern mining have been put forward (Das et al., 2020; Mowafy et al., 2018; Nasridinov et al., 2016). Most of them, however, have similar weaknesses. Either they extract patterns from a single source of crime data, or they do not consider any data sources. Crime data, nevertheless, come from different sources such as crime databases, news articles, and social media which provide data with varying amounts and forms which demand separate techniques for data collection, pre-processing, and analysis. In this regard, for example, a framework for structured data cannot be suitable for unstructured. Therefore, if the mining task involves both structured and unstructured data, it would mean that multiple frameworks are to be applied, which is inadequate and tedious. It was against this backdrop that the present study proposed a generic and flexible framework for extracting crime insights from diversified sources of data.

RELATED WORKS

Over the past years, scholars invested a great deal of research on the mining of crime patterns from datasets. This can be explained with the ever growing generation of data in recent years due an increasingly dependency on digital platforms. Accordingly, different frameworks for pattern mining have been introduced. A study by Mowafy et al. (2018) proposed a framework to predict the crime type from large amount of unstructured data from police incident reports. On the other hand, Keyvanpor et al. (2011) proposed a multipurpose framework for intelligent crime investigation. The framework uses data mining techniques to discover vital entities from police narrative reports. Likewise, Chen et al. (2004) proposed a crime pattern mining framework by integrating association rule learning techniques and knowledge discovery. This framework can identify different types of crimes from police department database.

Moreover, Iqbal et al. (2012) introduced a framework for mining forensic-related information from doubtful online messages. The Iqbal's framework goes through five stages as follows: capturing online messages, extracting social networks from the log, summarizing chat conversations into topics, identifying information relevant to crime investigation, and visualizing the knowledge generated. Further, a study by Nasridinov et al. (2016) suggested a framework for predicting crimes basing on data mining. The proposed framework was based on KNN, k-means, and skyline algorithms, and comprises the following four modules as follows: generation of test data, classification, clustering, and data ranking. Unlike others, the Nasridinov's framework does not consider any input data.

As described in the presented previous related works, most prevailing frameworks give on to similar challenges. One of which is input data consideration. Majority of the previous frameworks either consider only one source of input data (e.g., police incident reports (Mowafy et al., 2018), crime databases in police departments (Chen et al., 2004), chat logs (Iqbal et al., 2012), etc.), or do not consider any input source of data (Nasridinov et al., 2016). This is heeded a challenge for two grounds. First, patterns of crime can be extracted from more than one sources of data; therefore, consideration of input data variability is crucial. Second, the type of input data usually determines how the patterns extraction will be done. Thus, if the input data is unknown, it becomes hard to make decisions on the type of pattern extraction approach to be used. A second challenge regarding existing frameworks is lack of clear procedures for cleaning and pre-processing of the collected data.

METHODS

The study was based on systematic review of literature to establish the study gap followed by a comprehensive experimentation for the validation of the proposed framework. With regard to the systematic review, in roder to select the required literature the study employed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework. The literature that was used included journal articles, conference proceedings, reports, books, case studies, and informational web contents related to the focus of this study. To obtain the literature, a thorough search was conducted in various databases including DOAJ, AJOL, Google Scholar, JSTOR, EBSCO, and Web of Science. Searching was also done in other relevant sources. A total of 71 studies were identified, comprising 63 from databases and 8 from other sources. After removal of duplicate articles, 52 remained. Out of these, 18 were considered irrelevant and subsequently excluded. A detailed screening of the remaining articles led to the exclusion of 11 due to the inability to access their full texts. 23 full-text articles qualified for analysis; however, 6 were excluded because of technical issues. Consequently, the final count of studies included in the review was 17, as illustrated in the PRISMA framework in Figure 1.

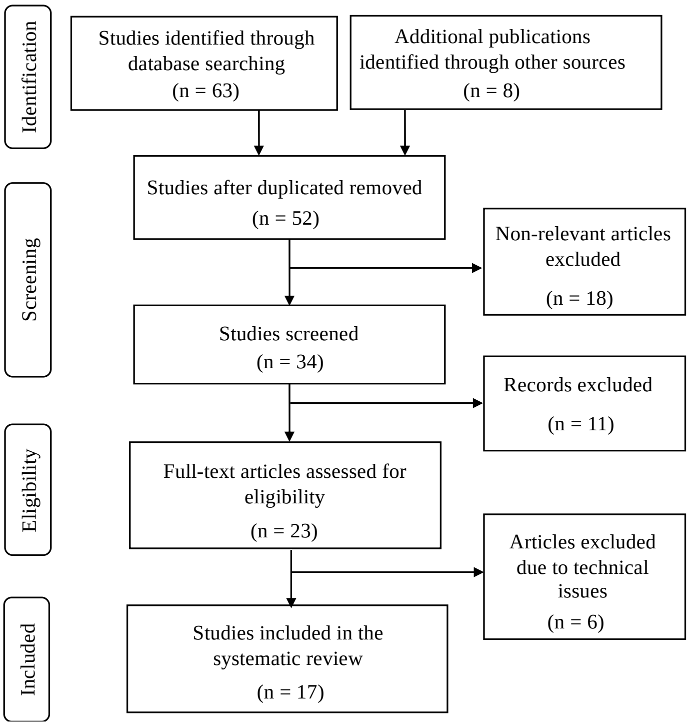

Figure 1: PRISMA framework showing inclusions and exclusions of reviewed articles

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

To accomplish experimentation of the suggested framework, two different sources of data were used. The first was crimes database which provided structured data, while the other one was news articles which provided unstructured data. Structured data was obtained from a crime database that was developed and populated with crime data. On the other hand, news articles were obtained from a Tanzanian online publishing media (i.e., IPP Media) available at: https://www.ippmedia.com/.

To run the experiments, a macOS Monterey (version 12.5.1) computer, M1 Pro with 16GB of memory was used. In order to obtain comparative results, the classical FP-Growth algorithm, along with the algorithm developed based on the proposed framework were employed. Java was used to develop a working prototype for experimentation purposes. The prototype was used in the extraction of patterns. In order to obtain reasonable results, structured datasets were divided into six different clusters of 500, 1000, 1500, 2000, 2500 and 3000 records while unstructured data was grouped into five clusters of 5MB, 10MB, 15MB, 20MB and 25MB. Each of these sets of data were analysed separately and the time to complete execution as well as amount of memory used were recorded.

RESULTS

The proposed Framework

To address the challenges identified in the existing frameworks, this work proposes a generic and flexible framework that extracts crime insights from diversified data sources. The proposed framework is made of four stages as follows: data sources, pre-processing, processing, and pattern visualization. Figure 2 presents the proposed framework.

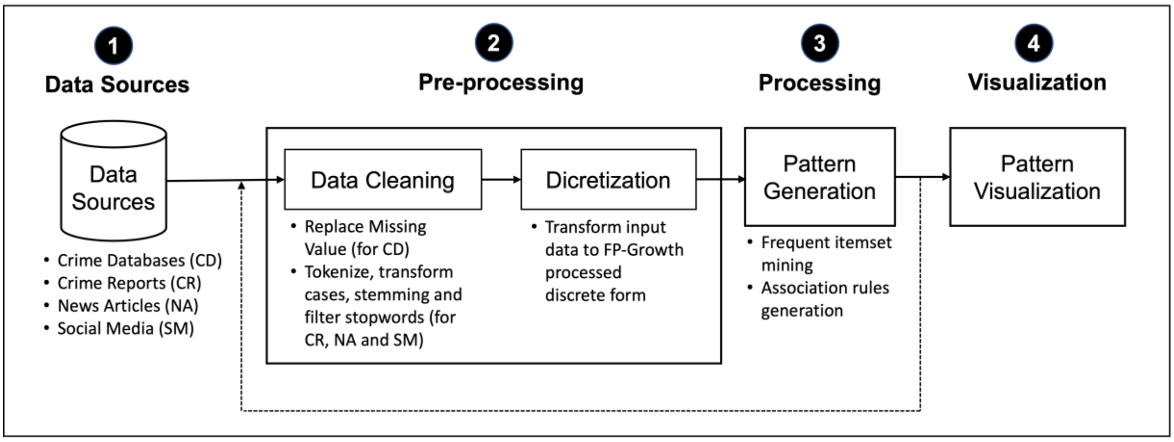

Figure 2: The Proposed Framework

Data Sources

Since crime patterns are extracted from available datasets, the framework puts data source as its first step. In this step data is collected from diverse sources for patterns extraction. The framework, however, focused on the following four crime data sources: crimes databases (CD), crime reports (CR), news articles (NA), and social media (SM). These sources were considered because several scholars, see for example Oghogho et al. (2024), Maslennikova et al. (2021), Chowdhury (2023), Mowafy et al. (2018), Matto and Mwangoka (2017), and Iqbal et al. (2012), use them as the main sources of crime data. Each of these sources offers a unique opportunity of availability of crime datasets, and thus stand to provide useful crime insights once analyzed. Table 1 describes each of these data sources in details. The presented data sources consist of datasets of different types. CD, for example, consists of structured data while SM consists of unstructured data. This therefore demands a careful consideration through a step-by-step approach during data pre-processing as explained further in the next section.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Table 1: Description of data sources employed in the proposed framework

Data Source	Description	Previous works that use the data source
Crimes databases (CD)	Electronic databases that contain comprehensive details regarding reported criminal activities such as the nature of crime, the methods employed, the timing and location of the offense, the identity of the perpetrator, and the underlying motivations. Additionally, it maintains an array of records pertaining to offenders, which includes their names, ages, genders, marital statuses, fingerprints, and photographs. These databases are held by the police, and once analyzed can provide useful crime insights.	Khatun et al. (2021); Carnaz et al. (2020); Matto and Mwangoka (2018); and Chen et al. (2004)
Crime incidents reports (CR)	Crime-related reports published by government and non-government agencies for public consumption. An analysis of a compilation of numerous such reports can provide valuable insights into criminal activities.	Maslennikova et al. (2021); and Mowafy et al. (2018)
News articles (NA)	Reports of crime incidents occurring in communities issued by various news media outlets, including newspapers, blogs, and informational websites. These reports can be examined to reveal valuable patterns and trends in criminal activities.	Prathap et al. (2021); Alatrista-Salas et al. (2020); Matto and Mwangoka (2017); and Iglesias et al. (2016)
Social media (SM)	User-generated content on online social networking platforms such as Facebook, Instagram, X, TikTok and WhatsApp that when analyzed, can help in providing useful apprehension with regard to crime.	Oghogho et al. (2024); Vo et al. (2020); and Iqbar et al. (2012)

Data pre-processing

After the data collection step, this step involves pre-processing of the collected data. Pre-processing entails preparation of the dataset for the actual processing. In the proposed framework, pre-processing involves two separate steps. The first one is data cleaning. Maharana et al. (2022) define data cleaning as the procedure of identifying and rectifying or eliminating erroneous, corrupted, improperly formatted, duplicate, or incomplete data present in the dataset. Since the dataset in this framework comes from diverse sources (i.e., structured and unstructured) data cleaning is based on the type of data. Structured data comes from CD in which missing values are common. Therefore, as part of the data cleaning, missing values are replaced with other values instead of ignoring or removing them. In the case of unstructured data (i.e., CR, NA and SM), missing values are not replaced because the data is naturally unstructured. Instead, the data cleaning involves the following steps: tokenization (words in the data file are grouped together and counted), transformation of cases (all contents in the data file is converted to lowercase or uppercase), and stemming (words are reduced into their stems or roots) and filter stopwords (stopwords are eliminated from the data file). Stopwords are the most frequently used words, including prepositions and pronouns, which are generally not beneficial for the data mining process.

Because the proposed framework is based on the FP-Growth algorithm, which works with discrete data, and because the input data are not necessarily discrete, discretization is second step in the data pre-processing. Discretization refers to the process of converting continuous data attributes into a limited set of discrete values, ensuring that the distinctions present in the original data are maintained within the resulting discretized dataset. It is identified as a fundamental contributor to association rules mining algorithms, as noted by Matto (2019). Different discretization methods exist. The proposed framework is based on the ChiMerge discretization method. ChiMerge is a supervised global discretization method, introduced by Kerber in 1992, that discretizes the data using Chi Square statistics (Kerber, 1992).

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Data Processing

This step involves the real processing of data to discover sets of items that occur frequently in a dataset. As previously explained, the frequent items are those that have a support equal to or exceeding the specified minimum support threshold (minsup). In other terms, an item X is considered frequent if support (X) ≥ minsup. The proposed framework is based on the FP-Growth algorithm to generate frequent patterns from the input datasets. FP-Growth compresses the input data into a prefix tree, as shown in the pseudocode of the algorithm in Figure 3. In this algorithm, the input data are the dataset (DB) and minsup (ξ), whereas the output is the FP-tree from which frequent itemsets are extracted.

```
Procedure: FPGrowth(DB, ξ)
Define and clear F-List: F[];
foreach Transaction T_i in DB do
        foreach Item a_j in T_i do
                F[a_i]++;
        end for
end for
Sort F[];
Define and clear the root of FP-tree : r;
foreach Transaction T_i in DB do
        Make T_i ordered according to F;
        Call ConstructTree(T_i, r);
end for
                    ⋮
```

Figure 3: Pseudocode for the FP-Growth algorithm

Input datasets for the proposed algorithm is crime data which, as said by Matto (2019), consist of itemsets that differ substantially in terms of frequency of occurrence. This is so because some crimes happen frequently, whereas others happen rarely. In this case, classical FP-Growth that discovers frequent itemsets by using a single minimum support is not appropriate. The proposed framework addresses the challenge by employing a multiple minimum support FP-Growth approach. The study adopted procedures and steps for obtaining multimple minimum support values from Matto and Mwangoka (2018). The data analysis step provides crime insights through patterns they generate. At this stage it is important to investigate the results obtained if they provide meaningful results. The evaluation result may mean to go back to the pre-processing step where further data cleaning and resetting of LS values can be done to obtain significant results. Upon being content with the results, the last step (i.e., pattern visualization) will be taken.

Patterns Visualization

Pattern visualization is the final step of the proposed framework. It involves presenting the discovered patterns in a form that makes them easily understandable to users. Visual aids such as graphs, cause and effect diagrams, and bar charts can help to improve the visualization of mined patterns. This step is a crucial component for law enforcement agencies and other involved parties to interpret the extracted patterns, enabling them to implement necessary measures related to crime prevention strategies.

Validation of the proposed Framework

In order to validate the proposed framework, experiments were conducted in which run time and memory use of the classical FP-Growth and the proposed approach were captured. Experimental findings regarding both runtime and memory utilization indicate the effectiveness of the proposed approach. On top of being effective in the extraction of patterns, this approach yields a better runtime and memory use than classical FP-growth. Figure 4 (a) and (b) illustrate the execution time of the proposed method in comparison to the FP-Growth algorithm (for similar sets of data) on crime datasets and news articles, respectively. Similarly, Figure 5 (a) and (b) show the memory use of the proposed approach against the FP-Growth algorithm on the crime datasets and news articles, respectively.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

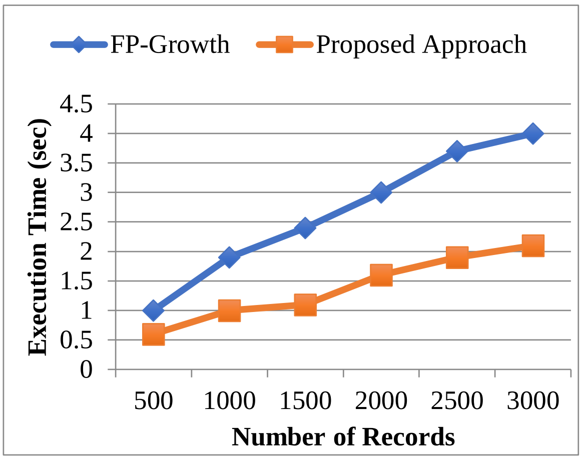

Figure 4(a): Run time on crime dataset

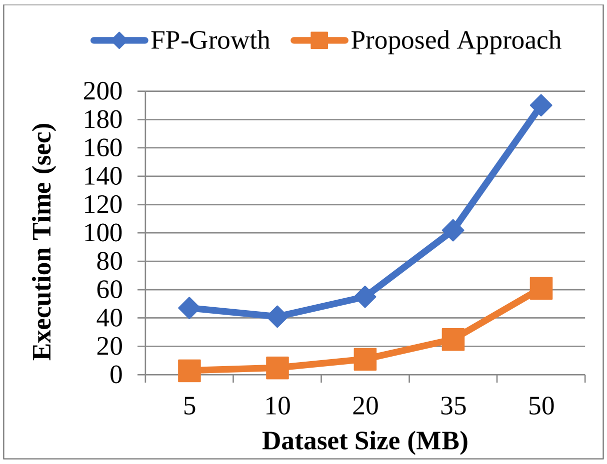

Figure 4(b): Run time on news articles

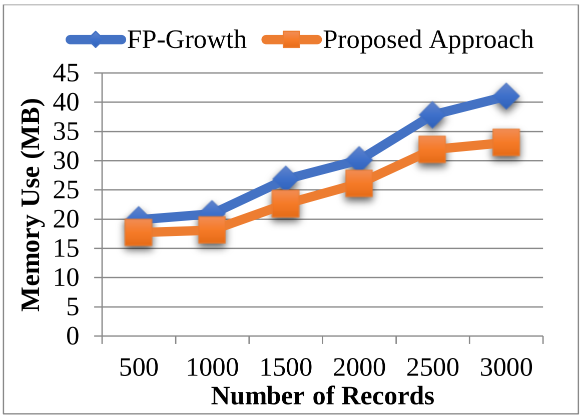

Figure 5(a): Memory use on crime dataset

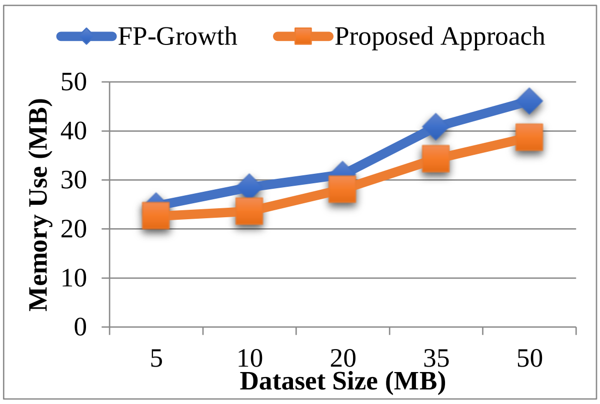

Figure 5(b): Memory use on news articles

CONCLUSION AND RECOMMENDATIONS

Multiple sources of crime data should be considered for effective extraction of crime insights. But for it to be effective, frequent pattern mining approach must consider using multiple minimum supports. In this research a comprehensive framework aimed at extracting crime patterns from diverse sources of crime data was introduced. The framework addresses the challenge of insufficient input data flexibility of the existing frameworks, provides clear guidelines on how to pre-process the input data, and considers the issue of rare items within crime datasets. It provides also flexibility in the mining of crime patterns derived from both structured and unstructured data. The framework is comprised of four stages. In the first step, the framework considers four data sources crime databases (CB), crime incidents reports (CR), news articles (NA), and social media (SM). The second step is concerned with data pre-processing. As part of pre-processing the framework suggests that two steps must be taken. The first one is data cleaning which based on the type of data. Data cleaning for structured data (i.e., CD) involves replacement of missing values while for unstructured data (i.e., CR, NA and SM) it involves tokenization, transformation of cases, stemming, and stopwords filtering. The second step in data pre-processing is discretization in which all types of data go through it. The third step is processing in which the real patterns are generated. And the last step is pattern visualization which involves presenting the discovered patterns in a form that makes them easily understandable to users. To assess effectiveness of the suggested framework, a series of experiments were carried out using a crime database and news articles as data sources. Experimental findings regarding both runtime and memory utilization indicate the effectiveness of the proposed approach for pattern mining. This proposed method demonstrates improved performance in terms of runtime and memory utilization when compared with FP-Growth. The study, nonetheless, recommends experimentation of the same method on crime reports and social media data to see if they will yield similar results.

Journal of Information Systems Engineering and Management
2025, 10(40s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

REFRENCES

[1] Agrawal, R. and Srikant, R. (1994). Fast Algorithms for Mining Association Rules. In: *Pro- ceeding of the 20th International Conference in Very Large Databases, VLDB*. Vol. 1215. pp. 487–499.

[2] Alatrista-Salas, H., Morzán-Samamé, J., & Nunez-del-Prado, M. (2020). Crime alert! crime typification in news based on text mining. In *Advances in Information and Communication: Proceedings of the 2019 Future of Information and Communication Conference (FICC), Volume 1* (pp. 725-741). Springer International Publishing.

[3] Carnaz, G., Beires Nogueira, V., Antunes, M., & Ferreira, N. (2020). An automated system for criminal police reports analysis. In *Proceedings of the Tenth International Conference on Soft Computing and Pattern Recognition (SoCPaR 2018) 10* (pp. 360-369). Springer International Publishing.

[4] Chen, H. Chung, W., Xu, J. J., Wang, G., Qin, Y. and Chau, M. (2004). "Crime data mining: A general framework and some examples," Computer, vol. 37, no. 4, 2004.

[5] Chowdhury, S. A. . (2023). Police Training for Managing Crime: With Special Reference to the Crime Data Management System of Bangladesh Police. *Social Science Review*, *40*(1), 73–88. https://doi.org/10.3329/ssr.v40i1.69076

[6] Das, P., Das, A. K., Nayak, J., & Pelusi, D. (2020). A framework for crime data analysis using relationship among named entities. *Neural Computing and Applications*, *32*(12), 7671-7689.

[7] De Farias, A. M. G., Cintra, M. E., Felix, A. C., & Cavalcante, D. L. (2018). Definition of strategies for crime prevention and combat using fuzzy clustering and formal concept analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *26*(03), 429-452.

[8] Dekker, R., van den Brink, P., & Meijer, A. (2020). Social media adoption in the police: Barriers and strategies. *Government Information Quarterly*, *37*(2), 101441.

[9] Fournier-Viger, P., Gan, W., Wu, Y., Nouioua, M., Song, W., Truong, T., & Duong, H. (2022, April). Pattern mining: Current challenges and opportunities. In *International Conference on Database Systems for Advanced Applications* (pp. 34-49). Cham: Springer International Publishing.

[10] Franchina, L., Sergiani, F., Brutti, G., & Donati, F. (2022). FP Growth Application for the Prediction of Terrorist Attacks. In *Proceedings of the Future Technologies Conference (FTC) 2021, Volume 1* (pp. 807-819). Springer International Publishing.

[11] Han, J., Pei, J., Yin, Y. and Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. *Data mining and knowledge discovery*. 8(1): 53–87.

[12] Iglesias, J. A., Tiemblo, A., Ledezma, A. and Sanchis, A. (2016). Web news mining in an evolving framework. Information Fusion 28 (2016) 90–98. Elsevier.

[13] Iqbal, F., Fung, B C. M. and Debbabi, M. (2012). Mining Criminal Networks from Chat Log. 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology.

[14] Kerber, R. (1992). "ChiMerge: discretization of numeric attributes," in Proceedings of the 9th National Conference on Artificial Intelligence—AAAI-92, pp. 123–128, AAAI Press, July 1992.

[15] Khatun, M. R., Ayon, S. I., Hossain, M. R., & Alam, M. J. (2021). Data mining technique to analyse and predict crime using crime categories and arrest records. *Indonesian Journal of Electrical Engineering and Computer Science*, *22*(2), 1052.

[16] Macingwane, A., & IsafIade, O. E. (2023, November). Investigating Frequent Pattern-Based Models for Improving Community Policing in South Africa. In *Southern African Conference for Artificial Intelligence Research* (pp. 203-218). Cham: Springer Nature Switzerland.

[17] Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, *3*(1), 91-99.

[18] Maslennikova, L., Vilkova, T., Sobenin, A., Tabolina, K., & Topilina, T. (2021). Using online services to report a crime. *Wisdom*, (2 (18)), 120-128.

[19] Matto, G. (2019). *Effective Mining of Crime Patterns from Growing Volumes of Data using Improved FP-Growth Algorithm*, Doctoral Dissertation, Nelson Mandela African Institution of Science and Technology.

[20] Matto, G. and Mwangoka, J. (2017). Detecting crime patterns from swahili newspapers us- ing text mining. *International Journal of Knowledge Engineering and Data Mining*. 4(2): 145–156.

[21] Matto, G., & Mwangoka, J. (2018). Mining Frequent Patterns of Crime using FP-Growth with Multiple Minimum Supports based on Shannon Entropy. *International Journal of Computer Applications*, 180(3), 45-52.
[22] Min, F., Zhang, Z. H., Zhai, W. J., & Shen, R. P. (2020). Frequent pattern discovery with tri-partition alphabets. *Information Sciences*, *507*, 715-732.
[23] Mokkadem, A., Pelletier, M., & Raimbault, L. (2023). Association rules and decision rules. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *16*(5), 411-435.
[24] Mowafy, M., Rezk, A., & El-bakry, H. M. (2018). General crime mining framework for unstructured crime data prediction. *International Journal of Computer Application*, *4*(8), 08-17.
[25] Mussa, A. M. (2023). Strategies used for Crime Prevention in Urban District, Zanzibar. *East African Journal of Education and Social Sciences,* 4 (1), 108 – 113.
[26] Nasridinov, A., Byun, J. Y., Um, N. and Shin, H. S. (2016). Application of Data Mining for Crime Analysis. Springer-Verlag Berlin Heidelberg. DOI 10.1007/978-3-662-47895-0_61.
[27] Nasyuha, A. H., Jama, J., Abdullah, R., Syahra, Y., Azhar, Z., Hutagalung, J., & Hasugian, B. S. (2020). Frequent pattern growth algorithm for maximizing display items. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *19*(2), 390-396.
[28] Oghogho, M. Y., Osazuwa, O. M. C., Ekeng-Ekeng, H., & John, I. G. (2024). Enhancing security in Nigeria: The implications of social media-based crime reporting. *The American Journal of Political Science Law and Criminology*, *6*(09), 22-39.
[29] Prathap, B. R., Krishna, A. V., & Balachandran, K. (2021). Crime analysis and forecasting on spatio temporal news feed data—an indian context. In *Artificial intelligence and blockchain for future cybersecurity applications* (pp. 307-327). Cham: Springer International Publishing.
[30] Roy, S., Bordoloi, R., Das, K. J., Kumar, S., & Muchahari, M. K. (2021, December). Association Rule Mining on Crime Pattern Mining. In *2021 International Conference on Computational Performance Evaluation (ComPE)* (pp. 269-272). IEEE.
[31] Sugiharti, L., Esquivias, M. A., Shaari, M. S., Agustin, L., & Rohmawati, H. (2022). Criminality and income inequality in Indonesia. *Social Sciences*, *11*(3), 142.
[32] Vo, T., Sharma, R., Kumar, R., Son, L. H., Pham, B. T., Tien Bui, D., Priyadarshini, I, Sarkar, M. & Le, T. (2020). Crime rate detection using social media of different crime locations and Twitter part-of-speech tagger with Brown clustering. *Journal of Intelligent & Fuzzy Systems*, *38*(4), 4287-4299.
[33] Zaki, M. J., Parthasarathy, S., Ogihara, M. and Li, W. (1997). New algorithms for fast discovery of association rules. In: *KDD*. Vol. 97. pp. 283–286.