International Journal of Advanced Technology in Engineering and Science
Vol. No. 09, Issue No. 04, April 2021
www.ijates.com
ijates
ISSN 2348 - 7550

REVIEW ON DATA MINING FRAMEWORKS: STRENGTH & WEAKNESS

Usman Dahiru Haruna

Department of Information Technology, ModibboAdama University of Technology, Yola, Nigeria.

Abstract

warehouse. Though some organization were able to achieve that and even further try to mine the data and understands pattern of relationship that exist therein. Yet crime is at its exponential increase because crime organization in developing nations fail to harmonize their data in to single store so as to be able to subject the data in data mining techniques and get matrix for making projection and prevention of future crime occurrences. In an effort to achieve a free crime society, this paper reviews some crime related data mining framework. but the analysis reveals that unstructured data is yet to be integrated and mine, the need to ensure quality data to subjected in to mining algorithms is equally important so as to good predicting variables for making inform decision making.

Keyword : Data, Data mining, Data mining technique, Data mining framework

1. Introduction

Data are been generated in and out of organizations in both public and private organizations. What is mind worrisome is that such organization doesn't know what do to with such data, only few of them are subject it to mining algorithms and discover the hidden pattern and knowledge hidden therein. Organizations like crime related organization gather data because it is needed for some operational purpose and ones it has served its purpose, it is left to languish on disk or tape or is discarded. Such data can be mined to discover hidden patterns or trends which can help improve the organization and even the society at large. These deficiencies have revealed the necessity of using a systematic and intelligent approach for crime investigation more than ever[1].Data mining automates the detection of relevant patterns in a database, using defined approaches and algorithms to look into current and historical data that can then be analyzed to predict future trends. Because data mining tools predict future trends and behaviors by reading through databases for hidden patterns, they allow organizations to make proactive, knowledge-driven decisions and answer questions that were previously too time-consuming to resolve. Data mining monitors each similar pattern from data gathered and helps in making prediction about the performance and also, enhance theoretical based research by giving its real name or numbers [2].

2.0 Data Mining

Data mining (DM) is one of the most powerful ways of knowledge extraction or we can say it is one of the best approaches to detect underlying relationships among data with the help of machine learning and artificial intelligence techniques [3]. Similarly [4] looked at data mining (sometimes called data or knowledge discovery) as the process of analyzing data from different perspectives and summarizing it into useful information to increase revenue, cuts costs, or both. But [5] looked at DM as a process of analyzing data from different perspectives and discovering imbalances, patterns and correlations in data sets that are insightful and useful for predicting results that help you make a good decision.

2.1 Data mining process

[6] Data mining process is not an easy process. It is complicated and has feedback loops which make it an iterative process. The data mining process involves six steps: (1) Problem definition, (2) Data Preparation, (3) Data Exploration, (4) Modeling, (5) Evaluation, and (6) Deployment while [5] present data mining process in the form of a model even though Ali also presented his own in the form of a model. But [5] model give a highly decouple and well elaborated view of the mining process as can be seen in fig. 1 below:

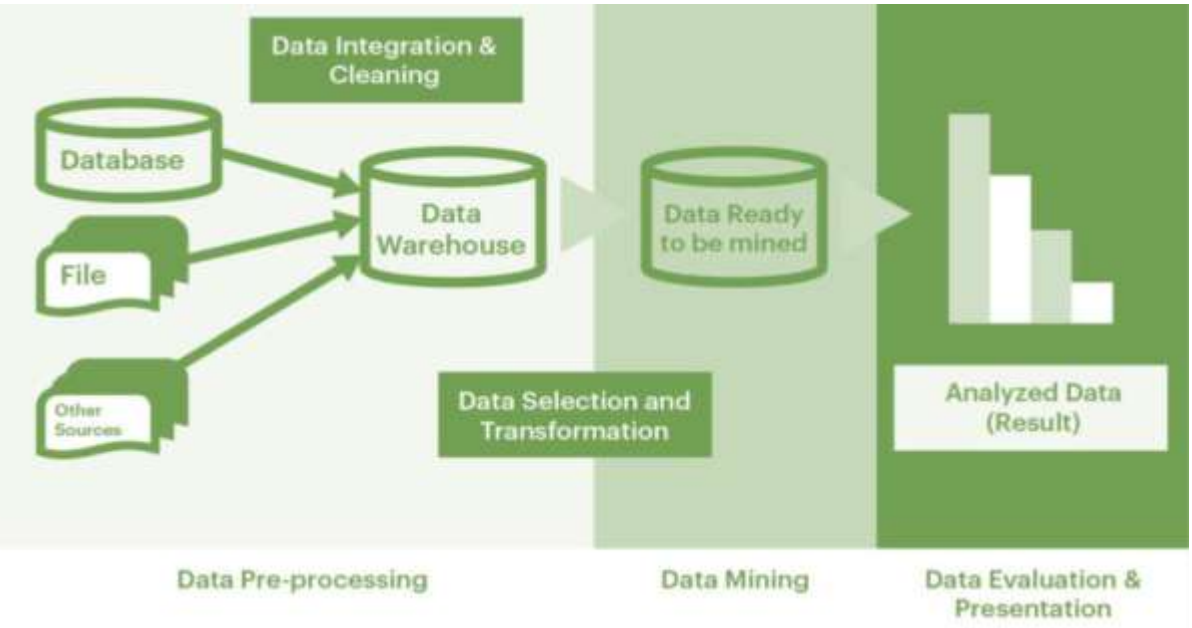

Figure 1:The Data Mining Process Model

Source: [5]

2.3 Data Mining Techniques

There are various data mining techniques that can be used to understudy the hidden pattern that exist on the existing repository and also identify the relationship between the dataset existing. [3] identified Entity extraction, Clustering techniques, Association rule mining, Sequential pattern mining, Deviation detection, Classification technique, String comparator, Social network analysis and Artificial neural networks as data mining techniques while [7] is of the view that Entity extraction, Clustering techniques and Association rule mining are techniques for mining. But [8] and [9] are of the opinion that several techniques are available but Artificial Neural Network is the best because it analyzes textual material best as compared to other techniques. However, [10] is of the view that the techniques can be summarized as can be seen in the fig. 2 below:

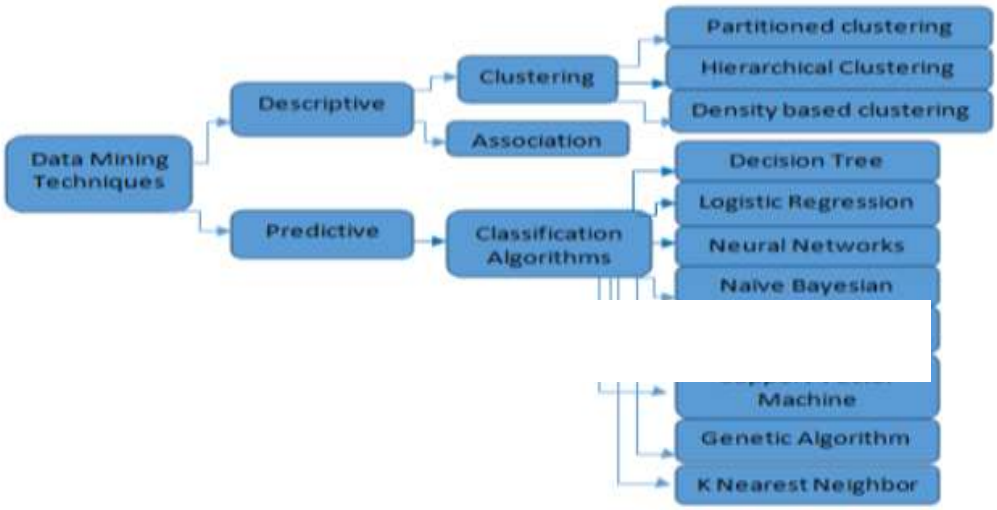

Figure 2: The Data Mining Techniques

Source: [10]

2.4 Review of Data Mining Frameworks

[11] build a model which will help predict the number of arrests that can be expected to be made for each crime, state-wise and age-group wise using regression. The major weakness of the model is proving that the data is not reliable enough to be used for framing new laws, preventing new crimes, bringing into place new strategies to curb these activities. Similarly, [12] build a crime predicting framework with clustering as a data analyzing technique in unsupervised data type. The framework provide security for the dataset during outsourcing. Clustering and

Classification is made on information while classifying the data, watermark content was used for verifying the classification data. The framework reveals that crime change over time and increase continuously, it identifies crime behavior and also manage large amount of crime data with certain degree of data security. The framework detects only hyper spherical clusters and it can't resolve the number of clusters in the crime data. this work asserts that, the result gotten by this framework is more reliable compared to [11] framework, because watermark was used to maintain a certain level of data integrity, though the framework will have used other mechanism to achieve high level of data integrity.

[13] comes up with a framework for predicting the frequency of several types of crimes and the frequency of anti-social behavior crimes. The model uses instance-based learning algorithm, Linear Regression and The M5P algorithm (is part of a category of algorithms called decision trees). The methodology adopted is CRISP-DM (Cross Industry Standard Process for Data Mining), which was found in comparison with other data mining methodologies to be well suited for predictive tasks for crime data. The results indicate that decision trees (M5P algorithm) can be used to reliably predict crime frequency in general, as well as anti-social behavior frequency. The framework was unable to handle time frame for prediction.

[14] when further to build a framework for the crime data analysis and detection using Classification, Association, Prediction and outliers and link analysis which helps specialists in discovering patterns and trends, making forecasts, finding relationships and possible explanations, mapping criminal networks and identify possible suspects. The framework predict crime as relate to age. Findings of this work shows classification rule was used to identify time and location as recommended by [13] and also pre-processing was adopted as separate entity to get clean accurate dataset which make the result of [11] framework unreliable.

[15] build a prediction framework using the scikit-learn Python Toolkit to predict crime type for improve decisions and reduce investigation efforts. It applied unstructured text classification model which depends on combining the Multinomial Naive Bayes as a classification technique, and TF-IDF as a vector space model for text extraction for more accurate text classification results. The framework was able to predict the crime type but cannot provide crime hotspot and can't predict the next expected crime region. Yet there is quite a significant improvement from the other research because the framework uses unstructured dataset, but unable to pin point on the region or hotspot and next crime region as can be seen from other frameworks.

[16] designed a framework which used optimized k-means algorithm for the clustering tasks and SVM to predict where next crime will happen in the form of clusters mapped on a boxplot presented in form of a graph. Prediction was achieved using structure data set but time was unable to be predicted while [17] designed mining prediction framework using deep neural network (DNN). which facilitated accurate and effective prediction of crime occurrences. It works well on high dimensional and multi-model data. But prediction cannot be applied on insufficient data and is unable to provide information regarding a specific crime type at a given time slot.

[18] come up with a framework for Crime Prediction and Forecasting in Tamilnadu using K-Means clustering, Agglomerative clustering and Density Based Spatial Clustering with Noise (DBSCAN) algorithms are used to

cluster crime activities based on some predefined cases. The result of K-Means clustering algorithm is visualized using Google Map. The K-Nearest Neighbor (KNN) classification is used for crime prediction. The performance of each clustering algorithms is evaluated using the metrics such as precision, recall and F-measure, and the results are compared. The drawback of the framework is that, agglomerative clustering will never redo any step i.e. once attached object can never be separated and criminal were not identified. [7] come up with a framework for crime data analysis. The framework uses machine learning. The data source for the framework is Law enforcement agencies collected data from different sources like telephone records, based social networks like Facebook, twitter, blogs, surveillance record, police records and financial transaction data for the crime investigation process.

CRISP-DM methodology was use for the framework which stands for Cross industry standard process for data mining. Criminal data are available in different formats, thus tackling the variety of data formats from multiple data sources and transforming the data into a desirable form became ineffective which make it not possible to predict crime hot spot and manage huge dataset.

3.0 Conclusion

Crime is a serious problem which should be undertaken and controlled by community as well as the whole world. Data that will be use to find lasting solution in such situation are there only few researches have considered and utilized such record that are left on silos data bases. Yet unstructured data set are yet not been considered and matrix for decision are not defined which will help in controlling and curbing crime. Data is the future crude oil and crime prediction and finding relevant information from large amount of crime data is very important.

4.0 Recommendations

This paper titled a review on data mining frameworks: strength and weakness recommend:

1. The full adoption and implementation of some these frameworks as an integrated enterprise software. for evaluation purpose and success quantifiable matrix visibility.
2. Further improvement should be done on the framework to accommodate Misspelling and grammatical mistakes that exist in crime data that exist within input parameters for the framework.

References

[1] N, Mahmud, A. Zinnah, R. Khalid &N. Ahmed,Crime cast: a crime prediction and strategy direction service.*International journal on Computer and Information Technology 18(2)*, 2016, 18-20.

[2] W. Pallavi, and R.Shelke. A study of data mining tools in knowledge discovery process. *International Journal for Research in Applied Science & Engineering Technology, 4(III)*, 2016, 74-79.

[3] Q. Shamailaand S. Hafsa. A survey of data mining techniques for crime detection university of sindh. *Journal of Information and Communication Technology, 2(1),* 2018 1-6.

[4] M. Savita andK. Gurjit. Introduction to data mining and data warehouse. *International Journal of Advanced Research in Computer Science,8(4),* 2017, 398-400.

[5] R. Renas and M. Revink. The Concept of Data Mining and Knowledge Extraction Techniques.*Qubahan Academic journal, 2(3),* 2020, 17-20.

[6] R. Ali and C. Bach, Data mining and warehousing. *American Journal of Engineering,3(4),* 2014, 34-39.

[7] T. DeepikaandS. Sanjiv, An approach to crime data analysis: a systematic review. International Journal of Engineering Technologies and Management Research, 5(2), 2018, 67-74.

[8] G. Rajkot,Data mining with neural network: a perspective.*International Journal of Advance Research in Engineering, Science & Technology, (Special Issue),* 2016 1-5.

[9] K. Harsh,N. Bharath, C.SiddeshandS. Kuldeep, An introduction to artificial neural network. *International Journal of Advance Research and Innovative Ideas in Education, 1(5),* 2017, 27-30

[10] R. Pallavi, C. Mandakinina and C. Radhika,A review on data mining techniques and challenges in medical field. *International Journal of Engineering Research & Technology*. 9(8), 2020, 329-333.

[11] R. Devakunchari, S. Bhowmick, S. BhutadaandY. Shishodia, Analysis of crimes against women in india using regression. *International Journal of Engineering and Advanced Technology, 8(2S2),* 2019, 124-127.

[12] V. VishnupriyaandM. Valarmathi,An effective data mining technique for analyzing crime patterns. *Journal of Computer Engineering, 1(7),* 2019, 26-30.

[13] S. Ginger andC.Mihaela, An exploration of crime prediction using data mining on open data. *International Journal of Information Technology & Decision Making, 16(05),* 2017, 1-27.

[14] B. SwadiandH. Fatlawi, Crime data analysis using data mining techniques to improve crimes prevention procedures. *Iraqi conference for Information technology, 1(3),* 2016, 34-39.

[15] M. Mona, A.RezkandH. El-bakry, Building unstructured crime data prediction model: practical approach. *International Journal of Computer Application, 8(4),* 2018, 1-7.

[16] M. Vijayalakshmi,B. ShivanshandG. Himanshu, Crime pattern recognition and prediction using optimsed k-means and svm. *International Journal of Pure and Applied Mathematics, 118(22),* 2018, 581-586.

[17] H. KangandB. Kang, Prediction of crime occurrence from multi-modal data using deep learning. *Journal of Pone, 12(2),* 2017, 44-48.

[18] S. Sivaranjani,S. SivakumariandM. Aasha, Crime prediction and forecasting in tamilnadu using clustering approaches. *IEEE International Conference on Emerging Technological Trends, 978*(1), 2018, 13-19.