



“Crime Software To Predict Places With High Crime Possibilities Using Machine Learning”

Sinchana B C¹, H P Mohan Kumar²

Research Scholar, Dept. of MCA, P.E.S College of Engineering, Mandya, India¹

Professor, Dept. of Computer Science & Engineering, P.E.S College of Engineering, Mandya, India²

Abstract: Now day's crimes are increasing rapidly and there is no region or city without crimes. So we require an automation for crime detection and prevention. System should stop crimes before it starts. As crimes are increasing, precautionary measures to be taken to stop crimes [2]. Finding frequent crimes and related correlations is a tedious and high effort task in the current crime sector. This issues needs to be addressed. Current real time system does not supports automation for crime prediction in real time. Machine learning or AI is the emerging technology to solve this issue. Efficient Unsupervised learning algorithms used to process crime training datasets and frequent crimes and their relationships are identified [5]. Proposed system build to reduce crimes and applicable to crime sector. We are building a real time application where it is useful for crime sectors to reduce the crimes. Currently none of the crime sector applications does this and we use efficient data science algorithms to predict crimes with better results.

Keywords: Data Science, Machine Learning, Apriori Algorithm, Apriori TID Algorithm, Crime Types, Patterns

I. INTRODUCTION

In real time crimes are increasing more and there are many different types of crimes such as pickpocket, murder, kidnap etc.... [1] Several crime related works have proposed various techniques to solve the crimes that used too many applications [3]. Crimes are increasing more in day today life. All cities and all over world facing this crime problems. Crime type depends on many factors such as location, population, school zone, hospital zone etc...[12] It is very important to identify the frequent crimes and take necessary actions to reduce the crimes. In this proposed work we build real time application for crime sector to find frequent crimes in an area and also we predict relationship between different types of crimes [8]. We apply efficient ML algorithms to process crime data and prediction is done. Data Science, is the trending technology that can applied to solve all type of problems in all types of fields [1]. In this work data science applied to crime sector to process old crime data or previous years crime data and to find the frequent crime types and their correlations [15].

Machine learning is used to train the system using training datasets. Here we use crime datasets, efficient machine learning algorithms used to process crime datasets and hidden crime patterns are extracted. Unsupervised learning techniques used to train the system and results predicted. Efficient unsupervised learning algorithms such as Apriori algorithm[5], Apriori TID algorithms[6] used to process crime datasets and frequent crimes and their relationships predicted. We develop a real time web application to predict frequent crimes and crime patterns using ML algorithms [11].

We build machine learning algorithm for predicting frequent crimes and their relationships and comparing the proposed algorithm with existing algorithm [13][14]. Our system does identifying crimes and related correlations and provides solution to reduce crimes and making public more alert and active. System finds the correlations between different crime types such as murder, chain snatching, kidnap, pick pocket, robbery etc.... We are building a real time application to find the frequent crimes happening today where it is useful for crime sectors to reduce the crimes.

II. RELATED WORKS

2.1 Survey Papers

2.1.1 IEEE PAPER TITLE: Crime Pattern Detection using Simple Apriori Algorithm.

YEAR OF PUBLICATION: 2020

AUTHORS: Peng Chenn, Justin Kurland.

METHODOLOGY: Apriori Algorithm Used.



DESCRIPTION: This paper aims at applying apriori algorithm to process crime data and predict crime patterns. Data science techniques are powerful subject to process data [9]. This paper aims to solve the crime problems and bring some solution using previously underutilized parameters from police recorded crime data [19]. In order to achieve the goal, a crime procedure is proposed and 3 factor (1) time; (2) setting; and (3) modus operandi.

LIMITATIONS:

- Uses efficient data mining algorithms.
- Huge data required.
- More time required for prediction.
- Not Implemented as Real Time A

2.1.2 IEEE PAPER TITLE: A Model for Visual and Intuitive Crime Investigation Based on Associative Rule Mining Technique

YEAR OF PUBLICATION: 2020

AUTHORS: Edigar ADERO, George OKEYO, Waweru MWANGI.

METHODOLOGY: Apriori Algorithm used.

DESCRIPTION: Crime has been part of our society ever since the concept of laws was first approved. There is an increased concern at governance level due to escalating levels of crime both internationally as well as locally in Kenya [11]. In this article, the researcher suggests the use of Associative Rule Mining to come up with a model suitable for crime analysis and prevention using Apriori algorithm to represent mutual implications among criminal occurrences [20].

LIMITATIONS:

- Apriori Algorithm used which process huge amount of crime data. Needs more data.
- Needs more time for crime data analysis.
- Doesn't predicts the crime relationships.

2.1.3 IEEE PAPER TITLE: Crime Forecasting Using Data Mining Techniques

YEAR OF PUBLICATION: 2020

AUTHORS: Chung-Hsien Yu, Max W. Ward, Melissa Morabito, and Wei Ding

METHODOLOGY: Classification Techniques Used.

DESCRIPTION: Most crimes are "undetermined." It is not necessarily random, but it also does not occur regularly in space or time [4]. In this research, we discuss preliminary results of a crime prediction model developed in collaboration with local police departments in the northern United States [12]. We analyzed several classifications to determine which method is best at predicting crime "hot spots." We also looked at the addition or formation of groups.

LIMITATIONS:

- Data mining Algorithms used, so huge amount of data required.
- Less Accurate Results.

2.2 Difference between Existing Works and Proposed Work

❖ In most of the current projects the implementation is successful, but the algorithms used are not programmed and they use off-the-shelf libraries and tools that the algorithms use. However, in the system concept, we program the algorithm, that is, we program own logic for the algorithm and the results are tested[7].

❖ Many research works uses less amount of training data-sets, in the proposed system we use huge data-sets for processing.

❖ All existing works uses PYTHON or R Language or Ready Data science tools for prediction and which works for static datasets, but in the proposed system we implement the concept for dynamic datasets (real time application)[10][14].

❖ All existing works are just model development, can't be used in real time. Here we build this concept as real time application using front end technology as "visual Studio" and back end technology as "SQL Server" and C# as programming language.

❖ Proposed system is a real world application with model using Microsoft technologies



III. CURRENT WORK

The proposed procedure can be applied to criminal cases. The planning process includes crime modelling, accurate detection, equipment planning and replacement, and operation time to find the necessary algorithms for crime detection. The proposed system detects criminal behavior, predicts crime, accurately analyzes and manages large amounts of information obtained from various sources. The proposed system uses data science technology “association rules” to predict the relationship between different types of crimes.

Example: location “vijaynagar”, depending on the previous crime data we can predict the different types of crime that takes place in future.

Output – Crime patterns with different crime types in “vijaynagar” for upcoming days.

1. robbery, murder related to suicide
2. child abusing related to murder
3. chain snatching, robbery related to murder
4. pick pocket, chain snatching related to robbery and murder

3.1 Datasets

| A | |
|----|--|
| 1 | Crimes |
| 2 | Murder, Kidnap, Chain Snatching, Robbery, Half Murder |
| 3 | Kidnap, Chain Snatching, PickPocket, Neck Hanging |
| 4 | Murder, Single Car Chain Snatching, Half Murder |
| 5 | Murder, Kidnap, Chain Snatching, Hit n Run |
| 6 | Hit n run, Drunk n Drive, Chain Snatching, Theft |
| 7 | Murder, Kidnap, Chain Snatching, Neck Hanging, Man Handling |
| 8 | Murder, Chain Snatching, Robbery |
| 9 | Drunk n Driven, Kidnap, Chain Snatching, Theft |
| 10 | Kidnap, Hit n run, Half Murder |
| 11 | Kidnap, Chain Snatching, Robbery, Half Murder |
| 12 | Murder, Kidnap, Chain Snatching, Robbery, Half Murder |
| 13 | Chain Snatching, Kidnap, Hit n run, Neck Hanging, Man Handling |
| 14 | Single Car Accident, Chain Snatching, Hit n Run |
| 15 | Bad Light, Kidnap, Chain Snatching, Hit n Run, Neck Hanging |
| 16 | Kidnap, Chain Snatching, Neck Hanging, Hit n Run |
| 17 | Murder, Single Car Chain Snatching, Half Murder |
| 18 | Murder, Kidnap, Chain Snatching, Hit n Run |
| 19 | Hit n run, Kidnap, Chain Snatching, Neck Hanging |
| 20 | Bad Light, Kidnap, Chain Snatching, Hit n Run, Neck Hanging |

Fig 1: Crime Datasets

IV. METHODOLOGY

4.1 Unsupervised Learning

A Descriptive model is used for tasks that would benefit from the insight gained from summarizing data in new and interesting ways. There are no predefined labels in unsupervised learning technique. The goal is to explore the data and find some structure with in. Unsupervised learning works well on transactional data. In our project Association learning algorithms used such as “**Apriori Algorithm and Apriori TID Algorithm**”. These algorithms preferred as algorithms supports small datasets and also large datasets.

Process Flow

Step 1: previous year’s crime data collected from different sources such as kaggle.com, dataworld.com, data.gov.in, github.com etc..

Step 2: here crime data is preprocessed, where we remove unwanted data and fetch the required data. Unwanted data such as crime no, serial no etc. will be removed.

Step 3: required data is inputted to algorithms. We use efficient algorithms such as Apriori algorithm and Eclat algorithm to process the data.

Step 4: after processing, frequent crimes will be extracted and displayed.

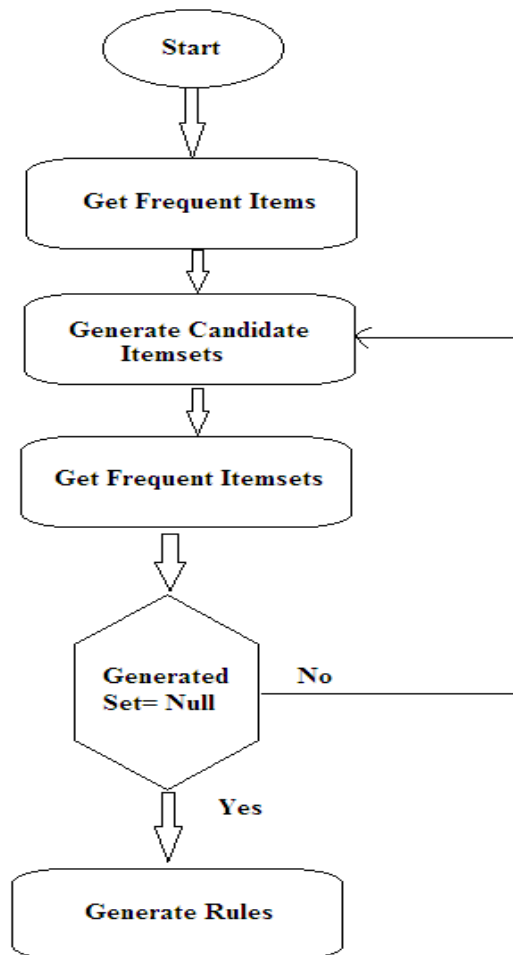


- Step 5: Relationship between different types of crimes will be extracted and displayed.
- Step 6: both algorithms results compared and efficient algorithm will be chosen.
- Step 7. Final outputs displayed on GUI.
- Step 8: Comparative analysis of algorithms displayed using graphs.

4.2 Apriori Algorithm Pseudo-code

```

Apriori (T, minSupport)
C1 = {candidate 1-itemsets};
L1 = {c ∈ C1 | c.count ≥ minsup};
FOR (k=2 ; Lk-1 > ∅ ; k++) DO BEGIN
Ck=apriori-gen(Lk-1);
FOR all transactions t ∈ D DO BEGIN
Ct=subset (Ck,t);
FOR all candidates c ∈ Ct DO
c.count++;
END
Lk={c ∈ Ck | c.count > minsup}
END
Answer=* Lk;
    
```



Flow of the Algorithm

Fig 2: Apriori Algorithm Flow



V. EXPERIMENT RESULTS

Graphical Representation – Frequency Plot – Jaynagar

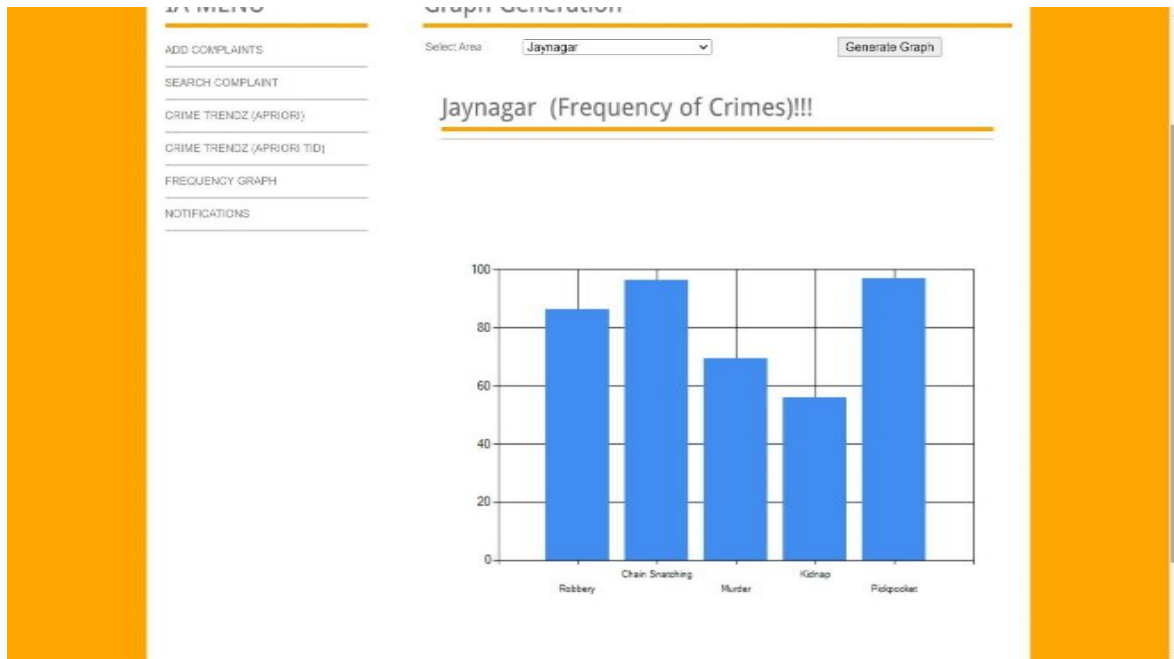


Fig 6: Crimes Frequency Plot

According to the crime frequency detection graph, robbery accounts for 88% of all crimes. Chain snatching tops the list with a staggering 94% frequency rate. Murder cases make up 69% of the total crimes reported. Kidnapping incidents account for 55% of the overall crime frequency. Pickpocketing has the highest frequency rate at 96%, making it the most common crime. Is all crime frequency detected as per the crime patterns project

VI. CONCLUSION

Identifying different types of crimes that can take place in upcoming days plays an important role in the current crime or investigation sector. Taking precautionary measures to avoid crimes is also important to reduce crimes in a city. Proposed system predicts frequent crimes and their relationship in advance, so that police or investigation departments can take necessary actions to reduce crimes. We are building a real time application where it is useful for crime sectors to decrease the number of crimes.

REFERENCES

- [1] S. Sathyadevan, M. Devan, and S. Surya Gangadharan, "Crime analysis and prediction using data mining," in Networks Soft Computing (ICNSC), 2014 First International Conference on, Aug 2014, pp. 406–412.
- [2] T. Pang-Ning, S. Michael, and K. Vipin, Introduction to Data Mining, 1st ed. Pearson, 5 2005.
- [3] S. Kaza, Y. Wang, and H. Chen, "Suspect vehicle identification for border safety with modified mutual information," in Proceedings of the 4th IEEE International Conference on Intelligence and Security Informatics, ser. ISI'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 308–318.
- [4] V. Vaithyanathan, K. Rajeswari, R. Phalnikar, and S. Tonge, "Improved apriori algorithm based on selection criterion," in Computational Intelligence Computing Research (ICCR), 2012 IEEE International Conference on, Dec 2012, pp. 1–4.
- [5] C. Chu-xiang, S. Jian-jing, C. Bing, S. Chang-xing, and W. Yun-cheng, "An improvement apriori arithmetic based on rough set theory," in Circuits, Communications and System (PACCS), 2011 Third Pacific-Asia Conference on, July 2011, pp. 1–3.
- [6] S. Kaza, T. Wang, H. Gowda, and H. Chen, "Target vehicle identification for border safety using mutual information," in Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, Sept 2005, pp. 1141–1146.



- [7] W. Huang, M. Krneta, L. Lin, and J. Wu, "Association bundle – a new pattern for association analysis," in Data Mining Workshops, 2006. ICDM Workshops 2006. Sixth IEEE International Conference on, Dec 2006, pp. 601–605.
- [8] N. Sasaki, R. Nishimura, and Y. Suzuki, "Audiowatermarking based on association analysis," in Signal Processing, 2006 8th International Conference on, vol. 4, Nov 2006.
- [9] A. Ben Ayed, M. Ben Halima, and A. Alimi, "Survey on clustering methods: Towards fuzzy clustering for big data," in Soft Computing and Pattern Recognition (SoCPar), 2014 6th International Conference of, Aug 2014, pp. 331–336.
- [10] A. Thammano and P. Kesisung, "Enhancing k-means algorithm for solving classification problems," in Mechatronics and Automation (ICMA), 2013 IEEE International Conference on, Aug 2013, pp. 1652–1656.
- [11] Y. Zhao and G. Karypis, "Evaluation of hierarchical clustering algorithms for document datasets," in Proceedings of the Eleventh International Conference on Information and Knowledge Management, ser. CIKM '02. New York, NY, USA: ACM, 2002, pp. 515–524. [Online]. Available: <http://doi.acm.org/10.1145/584792.584877>
- [12] C.-N. Hsu, H.-S. Huang, and B.-H. Yang, "Global and componentwise extrapolation for accelerating data mining from large incomplete data sets with the em algorithm," in Data Mining, 2006. ICDM '06. Sixth International Conference on, Dec 2006, pp. 265–274.
- [13] X.-M. Zhao, Y. ming Cheung, and D.-S. Huang, "Microarray data analysis using rival penalized em algorithm in normal mixture models," in VLSI Design and Video Technology, 2005. Proceedings of 2005 IEEE International Workshop on, May 2005, pp. 129–132.
- [14] H. Chen, W. Chung, Y. Qin, M. Chau, J. J. Xu, G. Wang, R. Zheng, and H. Atabakhsh, "Crime data mining: An overview and case studies," in Proceedings of the 2003 Annual National Conference on Digital Government Research, ser. dg.o '03. Digital Government Society of North America, 2003, pp. 1–5. [Online]. Available: <http://dl.acm.org/citation.cfm?id=1123196.1123231>
- [15] R. Marikhu, J. Moonrinta, M. Ekpanyapong, M. Dailey, and S. Siddhichai, "Police eyes: Real world automated detection of traffic violations," in Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2013 10th International Conference on, May 2013, pp. 1–6.
- [16] W. Cheng, X. Ji, C. Han, and J. Xi, "The mining method of the road traffic illegal data based on rough sets and association rules," in Intelligent Computation Technology and Automation (ICICTA), 2010 International Conference on, vol. 3, May 2010, pp. 856–859.
- [17] U. Thongsatapornwatana and C. Chuenmanus, "Suspect vehicle detection using vehicle reputation with association analysis concept," in Tourism Informatics, ser. Intelligent Systems Reference Library, T. Matsuo, K. Hashimoto, and H. Iwamoto, Eds., vol. 90. Springer Berlin Heidelberg, 2015, pp. 151–164.
- [18] A. Kondaveeti, G. Runger, H. Liu, and J. Rowe, "Extracting geographic knowledge from sensor intervention data using spatial association rules," in Spatial Data Mining and Geographical Knowledge Services (ICSDM), 2011 IEEE International Conference on, June 2011, pp. 127–130.
- [19] C.-H. Yu, M. Ward, M. Morabito, and W. Ding, "Crime forecasting using data mining techniques," in Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, Dec 2011, pp. 779–786.
- [20] G. Yu, S. Shao, and B. Luo, "Mining crime data by using new similarity measure," in Genetic and Evolutionary Computing, 2008. WGEC '08. Second International Conference on, Sept 2008, pp. 389–392.