

Association Rule Mining Based Crime Analysis using Apriori Algorithm



Hamidah Jantan¹ Aina Zalikha Mohd Jamil²

¹Universiti Teknologi MARA, Malaysia, hamidahjtn@uitm.edu.my

²Universiti Teknologi MARA, Malaysia, ayena_aina@yahoo.com

ABSTRACT

Nowadays, criminal law enforcement is a crucial task due to the increasing of crime rates, limitation of manpower and lack of awareness from the local community. Historical data on crimes activities need to be analyzed in order to get the trend and pattern of crimes for future prevention actions. The aim of this article is to explore the relationship between the category of location and area of criminal records through extracting the patterns that frequently occur by applying Apriori algorithm from Association Rule Mining (ARM) method. CRISP-DM methodology is used to conduct this study that consists of business and data understanding, data preparation, modeling, evaluation and deployment phases. As a result, there are strong rules being created from the high support and confidence in modeling process where the generated rules would be considered as potential rules for pattern visualization in crime analysis. This study brings a high significance for effectiveness and efficiency strategy in criminal law enforcement and it can also be explored for other association rule mining methods for future work enhancement.

Key words: Association Rule Mining (ARM), Apriori Algorithm, Crime Analysis.

1. INTRODUCTION

Business analytic plays an important role for any organization that consists of statistical and analysis operation applied to the past business records in order to improve the decision making processes [1, 2]. Therefore, by implementing this approach, it can help the organizations to make better decision by analyzing huge amount of data visually. There are three levels of data analyses which are referring to descriptive, prescriptive and predictive. Descriptive analysis is an understanding of organization environment trends and causes of some occurrences such as reporting on certain trend analysis and the key significance technology area is visualization. Prescriptive analytic refers to analysis in order to decide their best performance in making decision of the organization. This analysis aims to optimize the performance of the system. The predictive analysis aims to determine what is likely to happen in the future using several methods such as classification, clustering, prediction and associations [2, 3].

This approach will give a new direction of decision making in an organization towards their future sustainability and competitive advantage.

Association rule mining (ARM) is a data mining technique that is used to search the probability in a set collection of items. The technique involves a process of finding patterns, associates and correlates the pattern by producing the association rules [4, 5]. ARM is an unsupervised data mining technique that used to demonstrate the hidden relationship between the co-occurrences of data items [6]. Besides that, there are several ARM algorithms that can be uses to generate the association rules such as Apriori, FP-tree, Continuous ARM, Rapid ARM, Generalized Disjunctive Association Rule (d-rules) Mining, Genetic Algorithm etc. Apriori algorithm is known as highly efficient in generating frequent item sets by involving two step processes to find large item sets first, and then scans the database to check the support count of corresponding item set. It uses main memory to store the discovered rules and efficient data structures to recover the frequency of each association pattern [7-8]. There is quite a number of applications use the discovered hidden association pattern from this approach in their decision making process such as for text categorization, student's performance evaluation, sales analysis, crime analysis, marketing analysis and etc. [7].

Every day, either in an hour, a minute perhaps a second, various crime activities are occur in our sounding. Besides that, the crime's index is also increasing every year, it shows that each entity in a community should play important role in preventing these activities. The vicinity crime historical data can be analyzed to evaluate the crime pattern trend and risk about the crimes. Thus, the discovered hidden pattern will help the responsible entity for prevention actions plan. They can predict the potential future criminal activity for the specific crimes. Crime analysis is a process of analyzing the data by using a set of qualitative and quantitative techniques that can be used for the police agencies and community [9]. It is also defined as a systematic study that relate to police issues by adaption of qualitative and quantitative data [10]. As discussed before, the aim of business analytic is to prepare the high quality of information to managers or decision makers for better future planning with informative facts. For those reasons, this study attempts to explore the potential of Apriori algorithm from ARM technique as a method for crime analysis.

This paper is organized in the following manner: related work on crime analysis, ARM applications and techniques has been discussed in the second section. The third section describes the research method. Then the fourth section is on the result analysis and discussions; and followed by the conclusion and future work in the last section.

2. RELATED WORKS

2.1 Crime Analysis

Crime analysis is a systematic study that relate to police issues by adaption of qualitative and quantitative method for the police agencies and community future planning [10]. The systematic study refers to the crime and disorder problems as related to police issues and focusing on socio-demographic, spatial and temporal factors as shown in Figure 1. The crime and disorder problem are related to the nature of incidents such as loud noise complaints, suspicious activity with problems and many others. The police related issues are referring to the need of staff and the area that needs the services of police. Therefore, the significance of socio demographic, spatial and temporal factors is needed for the purpose of crime analyses. By the use of qualitative and quantitative methods, it can help police to prevent crime which is by the criminal apprehension, crime and disorder reduction, and crime prevention. Therefore, the aims of crimes analysis are to help and support the operations of police departments in their daily tasks.

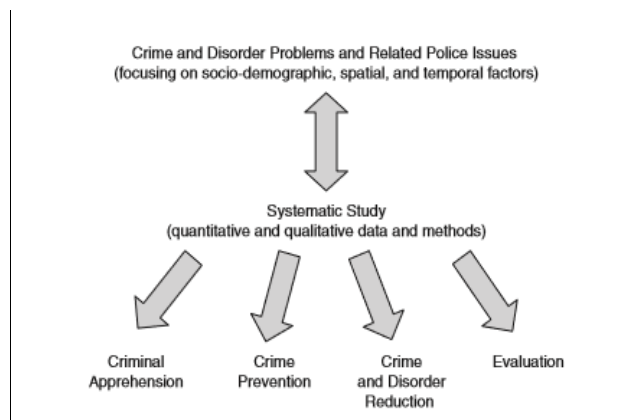


Figure 1: Crime Systematic Study (Source [9])

International Association of Crime Analysts has recognized four types of crime analysis which is crime intelligence analysis, tactical crime analysis, strategic crime analysis and administrative crime analysis [9, 10]. Business analytic processes supported that types of crime analysis where there are several factors need to be identified such as the purpose, scope, source of data and analysis techniques. Besides that, the strategic crime analysis is an analysis of data for long range problem and it is a must for a police department to have this kind of analysis for their future planning. Nowadays, the crime intelligence analysis is a new trend of crime analysis. This analysis uses intelligent methods in data mining

techniques for data analysis due to their abilities to produce highly accurate result of analysis. Table 1 shows several data mining techniques such as classification, prediction, association and clustering that have been used in crime analysis.

Table 1: Method and Techniques in Crime Analysis

Technique	Related studies	Description
Association Rule Mining	An Outlier-based Data Association Method for Linking Criminal Incidents. [4]	Apply outlier-based association method to a robbery dataset
Prediction (Linear Regression)	Using Data Mining Technique to Analyze Crime of Bangladesh. [11,12]	Used data mining techniques to forecast future crime trends of Bangladesh.
Classification and Clustering	Crime Classification and Criminal Psychology Analysis using Data Mining. [13]	This is research proposed on classify crime based on their level of seriousness and analyze psychology of murder by using clustering.
Clustering (K-Means algorithm)	Crimes Analysis using K-Means Clustering. [14,15]	Research on technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement.

2.2 Association Rule Mining

Association Rule Mining (ARM) is a data mining technique that used to search the probability of items in a collection of data or dataset. ARM is an unsupervised data mining technique that defined as a process of finding the patterns, associates and correlations of dataset in a database [5, 6]. Besides that, ARM is commonly used to search the relationship between the co-occurrences of any items in the datasets that are represented by their association rule. The generated association rule is representing the antecedent and consequent of the rule. However, the importance of rules is determined by the support and confidence. Table 2 show the basics understanding of rule assessment.

Table 2: Association Rule Description

Rule: $X \rightarrow Y$ [Supp (%), Conf (%)]	
X is the Left-Hand Side, LHS or the antecedent	Y called Right-Hand-Side, RHS or consequent
S is support denotes probability that contains probability that contains both X and Y	C is confidence denotes probability that a transaction containing X also contains Y.
$\text{Supp}(X \rightarrow Y) = \frac{\text{Number of basket that contains both}}{\text{Total number of baskets}} \times 100$	$\text{Conf}(X \rightarrow Y) = \frac{\text{Supp}(X, Y)}{\text{Supp}(Y)} \times 100$

The support and confidence are metric for association rules in constructing and influencing the association model. These metric are significance, influence and satisfied by producing the support and confidence comparing with the initial threshold value [5]. ARM algorithms consist of join-based algorithm, tree-based algorithm and pattern growth algorithm. There several ARM common algorithms used in previous studies such as Apriori algorithm, Eclat, FB-Growth and etc. Apriori Algorithms for discovering frequent Item sets that based on the principal /property that all sub sets of frequent item sets must be frequent. Eclat is based on constraint i.e. one which satisfies constraint is selected and binary integer programming, method is applied to solve. FP-Growth algorithm is a structure which is compact and stores quantitative information about frequent pattern in data base. It applies divide and conquer strategy [16, 17]. Apriori algorithm is the most popular due to influence’s algorithm ability especially for the Boolean association rules and the easy for parallelized [6].

2.3 Apriori Algorithm

Apriori algorithm is defined as the fast association rules algorithm which uses the main memory to store the discovered rules and efficient data structures to recover the frequency of each association pattern [6, 8]. Apriori algorithm mostly works on transactional database that each of the transaction contains the itemset which is a set of items. This algorithm is combined with two significant steps i.e.: Joining step: joining makes feasible combinations of item sets with their respective support count. Second stage is the Pruning step: scan the database and verify the support count value is higher than or equal to the user-defined support count value [18, 19]. There are many previous studies that use this algorithm in solving their problem such as in insurance industry, retailing and crime analysis. Table 3 shows several

related studies that use this algorithm. In crime analysis, this algorithm also attracts attention of researchers in this area due to the ability of this algorithm in producing highly related result for future decision making actions.

Table 3: Works Related with Apriori Algorithm

Title	Description
Market Basket Analysis in Insurance Industry[20]	Research in 300 data clients of insurance company by using demographic variable data which focusing on grouping customer and find hidden patterns.
Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transaction [21]	Prediction of product sales trends and customer behavior. By this analysis, retailer is provided the report based on prediction and allow the hands-on and knowledge driven decision.
Crime Analysis Based on Association Rules Using Apriori Algorithm [6]	The NIBRS crime dataset are the purpose study in this research to predict the unknown characteristics of a specific case by taking past criminal cases.
Usage of Apriori Algorithm of Data Mining as an Application to Grievous Crimes against Women [19]	This research discovers about underlying pattern involved in court’s record from various section and Molestation is a worldwide issue.
ARCA: Mining Crime Patterns Using Association Rules[8]	Research on simplicity and effectiveness of ARCA. The ARCA approach by using Apriori algorithm are recognized to prevent crime and based on results crime motivation and crime severity are the recommendation

3. RESEARCH METHOD

This study adapts the Cross-Industry Standard Process for Data Mining (CRISP-DM) model that consists of six main phases. Fig 2 show the research framework of this project based on CRIPS-DM model that involves the business understanding, data understanding, data preparation, modeling, evaluation and deployment. In preliminary study

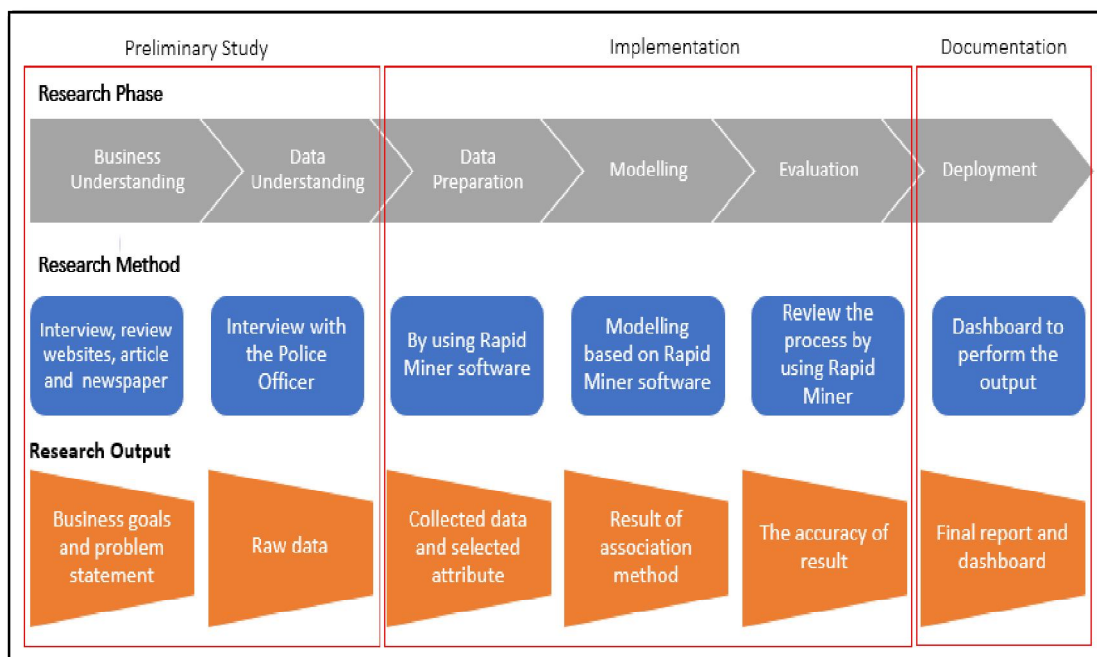


Figure 2: CRIPS-DM Based Crime Analysis Project Framework

business understanding is conducted to define the activities such as the background of study, business process, business problem by the interview session and literature study. The interview session was conducted to get better understanding on current situation, process flow, objectives of the organization and problems faced in the organization. In this study, several interview sessions were conducted with Police Officers from selected identified area for a case study.

The data understanding phase consists of the process of data collection, data exploration and quality of data verification. The collection of data is a process of gathering related data from selected Police department. 500 of crimes records were collected from the crime report and then stored using Microsoft Excel application. Table 4 shows the description of dataset used in this study.

Table 4: Crime Dataset Description

Attribute	Description	Data Type
Time	The time when the criminal activities occur in 24-hour time per day	Integer
Day	The days that represent 7 days per week	String
Type of Crime	Describe the type of criminal activities	String
Location	Describe the location that represent the crime case	String
Category of Location	Represent the category location of crime case	Varchar
Police Station	Represent the office or headquarters of a local police force	String

Next, the data exploration process was conducted to explore the crime dataset to analyze the data and information for further analysis. In this study, the Tableau software is used to explore the patterns of dataset as shown in Figure 3.

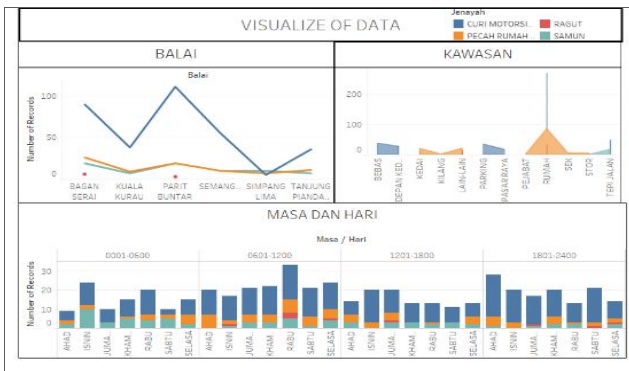


Figure 3: Crime Dataset Exploration

Then, the data quality verification process was performed to evaluate the quality of data by analyzing the dataset. This analysis is a crucial task because quality of data refers to ability of the dataset to serve an intended purpose of the analysis. For the case of high quality, it shows that the dataset is good for future analysis. In this study, Alteryx software is used to verify the quality of data for each year in each of selected attribute. Figure 4 shows the status of data quality

from one of the attributes in crime dataset.

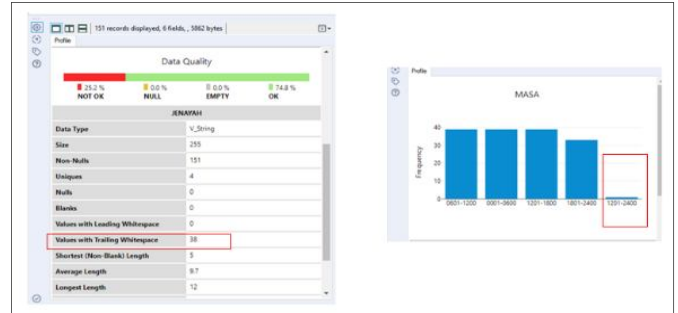


Figure 4: Crime Dataset Data Quality Verification

In data preparation phase, there are five stages were conducted i.e. select data, clean data, construct data, integrate data and format data using Alteryx software. Sample of the process of preparing clean data process is shown in Figure 5. Each item for each attribute in the dataset was checked by the frequency before proceeds to modeling phase.

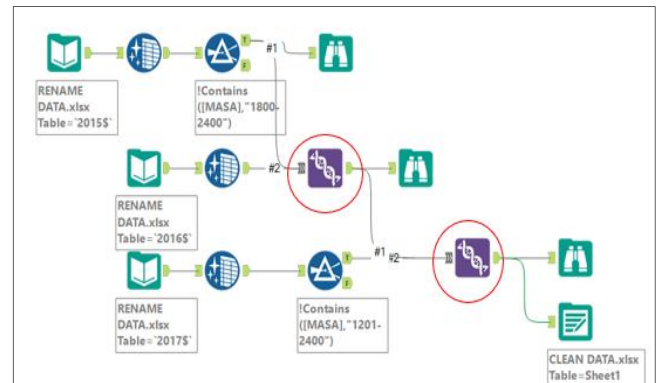


Figure 5: Clean Data Preparation Using Alteryx Software

In modeling phase, Apriori algorithm for association rules mining (ARM) is used to produce crime association rules. There are several pre modeling processes involved such as read, filter, aggregate, pivot, rename and replace the item set. Fig 6 shows the pre modeling process using Rapid miner software.

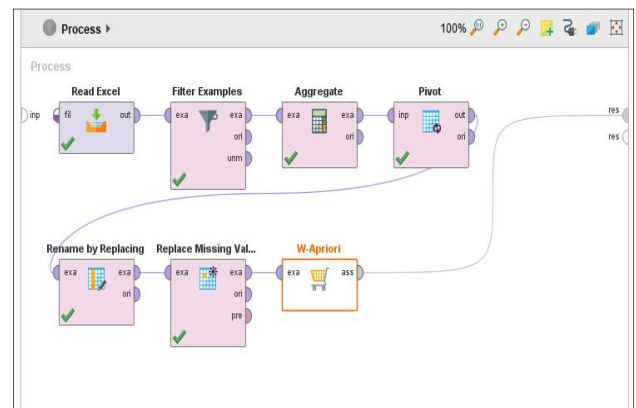


Figure 6: Modeling Crime Dataset using Rapid Miner Software

Figure 7 shows the ARM modeling process to generate the association rules. The minimum support and minimum confidence values are used to select the frequent items for strong rules generation. In evaluation phase, several experiments were conducted to evaluate the generated association rules by adjusting the threshold values of support and confidence.

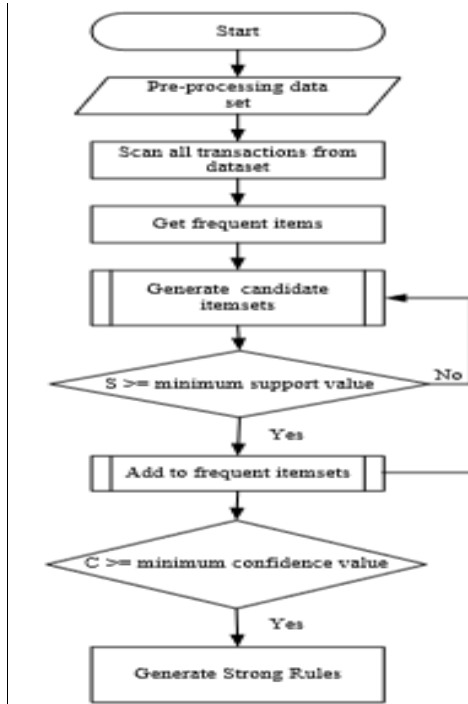


Figure 7: ARM Modeling Process of Data Analysis

In deployment phase, the dashboard for crime analysis was developed based on the evaluated association rules. The system will visualize some important aspects that need future action and propose some recommendation based on discovered pattern from the association rules mining process.

4. RESULT AND DISCUSSION

This study performed two experiments using selected crime dataset collected from police department. In business and data understanding phase, the result of analysis from visualization process was discovered several attribute or parameters that need future analysis. Sample of analysis is shown in Fig. 3 where the trend or pattern of selected crimes was discovered for future analysis and actions. Based on this analysis, there are two types of data set selected which are police Station category and days category. In experimental phase, the modeling and evaluation process was involved where Apriori algorithm is used to generate the association rules. In this study, the sample of association rules for crime analysis generated by Apriori algorithm is shown in Figure 8. The generated rules are depend on the support and confidence values as measurements to produce strong rules for association. The result of these experiments is shown in Table 5 where the analysis is based on different minimum support and confidence values for two different datasets.

```

W-Apriori

Apriori
-----

Minimum support: 0.25 (1 instances)
Minimum metric <confidence>: 0.2
Number of cycles performed: 15

Generated sets of large itemsets:

Size of set of large itemsets L(1): 6
Size of set of large itemsets L(2): 3

Best rules found:

1. 1=JLN SERINDIT , PARIT BUNTAR 1 ==> TEGORI TEMPT EJDI=RUMAH 1   conf:(1)
2. TEGORI TEMPT EJDI=RUMAH 1 ==> 1=JLN SERINDIT , PARIT BUNTAR 1   conf:(1)
3. 1=KLINIK JLN PASAR , PARIT BUNTAR 1 ==> TEGORI TEMPT EJDI=LAIN-LAIN 1   conf:(1)
4. TEGORI TEMPT EJDI=LAIN-LAIN 1 ==> 1=KLINIK JLN PASAR , PARIT BUNTAR 1   conf:(1)
5. 1=EDUTEX TRADING JLN BISTARI , PARIT BUNTAR 1 ==> TEGORI TEMPT EJDI=PEJABAT 1   conf:(1)
6. TEGORI TEMPT EJDI=PEJABAT 1 ==> 1=EDUTEX TRADING JLN BISTARI , PARIT BUNTAR 1   conf:(1)
  
```

Figure 8: Sample of Crime Association Rules

Besides that, the number of generated rules for each experiment was taken to determine the most significant rules for deployment phase. The result shows that dataset for days in second experiment created strong association rules that based on support and confidence values. Then, the generated association rules are used in development of application for crime analysis.

Table 5: Datasets Association Rules Analysis

Dataset	Min Support and Min Confidence	Rules	Result Min Support and Min Confidence	Conclusion
Dataset 1: Focused on Days	Support: 0.1, Confidence: 0.2	245 rules were created	Support: 0.33 Confidence: 1	↑
	Support: 0.4, Confidence: 0.8	37 rules were created	Support 0.4 Confidence: 1	=
Dataset 2: Focused on Police Station	Support: 0.1, Confidence: 0.2	245 rules were created	Support: 0.16 Confidence: 1	↑
	Support: 0.4, Confidence: 0.8	245 rules were created	No large itemset and rules found	0

The proposed system for crime analysis can visualize the trend of data based on selected attributes. As example, there is a navigation map where the user can choose to look for data analysis such as on police districts and crime analysis as shown in Figure 7, Figure 8 and Figure 9.



Figure 7: Navigation Map for Crime Analysis

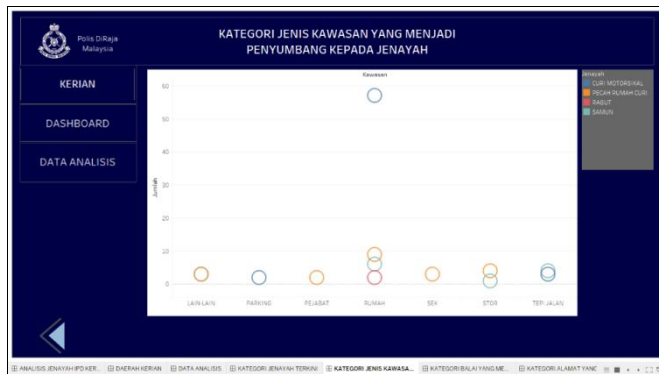


Figure 8: Sample of Crime Analysis

Besides that, the system can also proposed some recommendation to the police department based on the discovered knowledge as represented by the association rules as shown in Fig. 9.

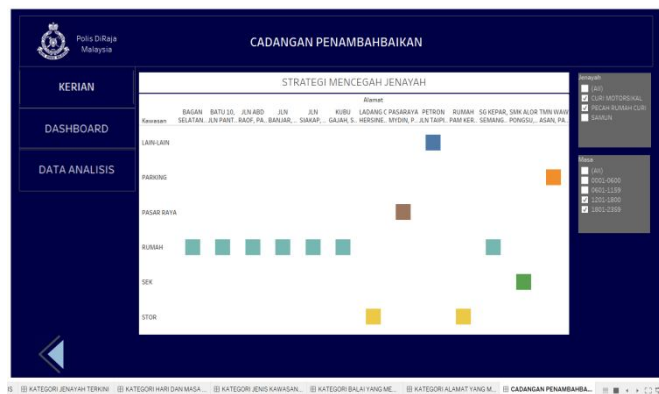


Figure 9: Recommendation from Crimes Analysis for Future Action

As example, if the police department needs to know about crimes in the specific area, they can identify the specific category of location and the frequency of crimes in that area. Based on association rules generated by the Apriori algorithm, the system can propose a suggestion to Police district for their future planning in order to reduce the crime’s rate for their district.

5. CONCLUSION

This paper proposed ARM based for crime analysis using Apriori algorithm. However, the result focuses on selected attributes for a case study for the specific type of crime and for a specific area. The results show that Apriori algorithm can produce highly significant knowledge from ARM for crime analysis. Besides that, there are other ARM algorithm such as FP Growth, Eclat and others algorithms need to be experimented as a comparative study to strengthen this findings. It would give a new understanding in this field especially to select the most significant result produced by the algorithm for crime analysis. Finally, this result will give a new direction of applying ARM algorithms in crime analysis for different crime datasets.

REFERENCES

- [1] Khan, R. A., & Quadri, S. M. K. (2012). Business Intelligence: An Integrated approach. *Business Intelligence Journal*, 5(1), 64-70.
- [2] Sharda, R, Delen. D & Turban E. (2014). *Business Intelligence and Analytics: System for Decision Support*. England. Pearson Education Limited.
- [3] Ranjan, J. (2009). Business intelligence: Concepts, Components, Techniques and Benefits. *Journal of Theoretical and Applied Information Technology*, 9(1), 60-70.
- [4] Lin, S., & Brown, D. E. (2006). An Outlier-based Data Association Method for Linking Criminal Incidents. *Decision Support Systems*, 41(3), 604-615. <https://doi.org/10.1016/j.dss.2004.06.005>
- [5] Dhayashankar, J. M., & Ramani, A. V. (2018). Pattern generation based Reverse engineering on Database Relation scheme using Enhanced Association rule mining. *International Journal of Applied Engineering Research*, 13(8), 6424-6432.
- [6] Sevri, M., Karacan, H., & Akcayol, M. A. (2017). Crime Analysis Based on Association Rules Using Apriori Algorithm. *International Journal of Information and Electronics Engineering*, 7(3), 99-107. <https://doi.org/10.18178/IJIEE.2017.7.3.669>
- [7] Mittal, K., Aggarwal, G., & Mahajan, P. (2017). A Comparative Study of Association Rule Mining Techniques and Predictive Mining Approaches for Association Classification. *International Journal of Advanced Research in Computer Science*, 8(9), 365-372. <https://doi.org/10.26483/ijarcs.v8i9.4984>
- [8] Pereira, B. L., & Brandão, W. C. (2014). ARCA: Mining Crime Patterns Using Association Rules. In 11th International Conference Applied Computing. Porto (pp. 159-165).
- [9] Elder, S., Bruce, C. W., Santos, R. B., Rodriguez, E., & Wyckoff, L. (2015). Definition and Types of Crime Analysis. *International Association Crime Analysis*. (IACA) Denver, Colorado.
- [10] Tripathi, S., Amendola, K. & Wyckoff, L. (2015). *Crime Analysis with Crime Mapping*. Crime Mapping and Analysis News: A Police Foundation Publication. Summer Issue (3).
- [11] Md. Abdul Awal, Jakaria Rabbi & Imran Rana. (2015). Using Data Mining Techniques to Analyze Crime of Bangladesh. *International Journal of Computer Science and Network*, 6(4), 489-494.
- [12] Elvis P.& Ronita E. (2019). Crime Prediction using Autoregressive Integrated Moving Average (ARIMA) Algorithm. *International Journal of Advanced Trends in Computer Science and Engineering*, 8(3), 720-724. <https://doi.org/10.30534/ijatcse/2019/59832019>
- [13] Samanta, A., & Joglekar, A. (2016). Crime Classification and Criminal Psychology Analysis using Data Mining. *International Journal of Mechanical Engineering and Information Technology*, 4(10), 1782-1789. <https://doi.org/10.18535/ijmeit/v4i10.01>

- [14] Agarwal, J., Nagpal, R., & Sehgal, R. (2013). Crime analysis using K-means clustering. *International Journal of Computer Applications*, 83(4).1-4.
<https://doi.org/10.5120/14433-2579>
- [15] Elvis P.& Ronita E. (2019). Crime Trend Analysis Using Data Mining Technique. *International Journal of Advanced Trends in Computer Science and Engineering*, 8(3).663-666.
<https://doi.org/10.30534/ijatcse/2019/52832019>
- [16] Supriyamenon, M. and Rajarajeswari, P. (2017). A Review on Association Rule Mining Techniques with Respect to their Privacy Preserving Capabilities. *International Journal of Applied Engineering Research*, 12(24). 15484-15488.
- [17] Sidhu, S., Meena, U. K., Nawani, A., Gupta, H., & Thakur, N. (2014). FP Growth algorithm implementation. *International Journal of Computer Applications*, 93(8).
<https://doi.org/10.5120/16233-5613>
- [18] Gaikwad, P. R., Kamble, S. D., Thakur, N. V., & Patharkar, A. S. Evaluation of Apriori Algorithm on Retail Market Transactional Database to get Frequent Itemsets. *Proceeding of Proceedings of the Second International Conference on Research in Intelligent and Computing in Engineering*.187–192.
- [19] Bansal, D., & Bhambhu, L. (2013). Usage of Apriori algorithm of data mining as an application to grievous crimes against women. *International Journal of Computer Trends and Technology*, 4(19), 3194-3199.
- [20] Roodpishi, M., & Nashtaei, R. (2015). Market basket analysis in insurance industry. *Management Science Letters*, 5(4), 393-400.
<https://doi.org/10.5267/j.msl.2015.2.004>
- [21] Prasad, P. & Malik, L. (2011). Using Association Rule Mining for Extracting Product Sales Patterns in Retail Store Transactions. *International Journal on Computer Science and Engineering*, 3(5), 2177-2182.