Journal of Applied Intelligent System (e-ISSN : 2502-9401 | p-ISSN : 2503-0493)
Vol. 6 No. 1, 2021, pp. 32 – 39

# Data Mining Applications for Violence Pattern Analysis with FP-Growth Algorithm

**Junta Zeniarja*[1], Debrina Luna Arghata Mangkawa[2], Abu Salam[3]**
*Dian Nuswantoro University, Imam Bonjol No. 207, Semarang, (024) 3517261*
*E-mail : junta@dsn.dinus.ac.id*[1], debrinalunaa@gmail.com[2], abu.salam@dsn.dinus.ac.id[3]*

**Abstract -** Violence is a criminal act which is one of the main problems based on each country. Violence can be defined as behavior that results in harm to a person. According to the research results of the Central Java Province DP3AKB (Office of Women's Empowerment, Child Protection and Family Planning) in 2017 there were approximately 200 people in Central Java province who experienced violence. Because of the violence that occurs with various forms of violence, it requires definite information about what kind of violence occurs most often. In obtaining this information, data mining techniques using the FP-Growth algorithm are needed. The application of the FP-Growth algorithm to produce an associated pattern of forms of violence. Hardness data 420 data, 7 best rules have been obtained with the support of min value 50% and min value support 60%. On the results of the best rules given recommendations (solutions) so that the DP3AKB can handle the problem of violence properly and on target.

**Keywords –** Data Mining, Associations, FP-Growth Algorithms, Hardness, Lift Ratio

## 1. INTRODUCTION

Today's developments have an impact on the development of society, for example people's behavior, people's mindset, as well as cultural shifts, coupled with an increase in population density which also results in high crime, one of which is violence. Violence is a criminal act which is one of the main problems experienced by every country and anyone can become a victim or perpetrator of violence. In general, violence can be defined as a behavior that causes other people to suffer and there is legal and normative rejection of this behavior [1]. There are various forms of violence, namely physical violence, psychological violence, verbal violence, sexual violence, exploitation, domestic violence, human trafficking and neglect.

According to research results from the Office of Women's Empowerment, Child Protection and Family Planning, Central Java Province (DP3AKB), there are many acts of violence experienced by all genders, both girls and adults as well as boys and adults. The total population in Central Java province is 34.26 million, in 2015 there were approximately 4600 people experiencing violence and in 2016 there were approximately 5300 people experiencing violence. This fact shows that there is an increase in the number of acts of violence. Usually, efforts that can be made to reduce the number of violence are counseling and seminars on violence in each city with the aim that the community understands the dangers of violence and understands various forms of violence, so that if someone experiences violence or sees violence, it is immediately reported to the parties concerned. However, this has not succeeded in reducing the amount of violence.

Because there are still many acts of violence with various forms of violence, it requires definite information about what forms of violence are most often violent, so that the government can find out the forms of violence that most often occur in each city and can then

follow up with discussions targeted topics of counseling and seminars, so that violence can be reduced. In obtaining this information, a data mining method is required.

Data mining is a series of processes to explore the added value of a data set in the form of knowledge that so far cannot be known manually [2]–[4]. The purpose of data mining is to find relationships or patterns that can provide benefits to related parties [5]–[7]. Data mining uses a discovery based approach, namely pattern-matching and algorithms used to determine key relationships in the data [8], [9]. The method used in this data mining management is Association Rules. Association Rules is a method that aims to find patterns that often appear in the data [10]. This method has several algorithms, one of which is the Frequent Pattern-Growth (FP-Growth) algorithm [10]–[19]. FP-Growth is an alternative algorithm that can be used to determine the most frequent itemset in a data set in the form of a prefix-tree structure or often called the FP-tree [12], [20], [21].

The FP-Growth Algorithm is used to find patterns of forms of violence. The FP-Growth algorithm has been widely applied, for example, in determining the relationship pattern of traffic accidents and its application to the library information system. The FP-Growth algorithm was chosen because it is very efficient in searching for frequent itemsets because this algorithm uses a different approach from the paradigm that has been frequently used, namely the a priori algorithm. This algorithm stores frequent itemsets in the form of an FP-Tree [22]. The FP-Tree that is formed can compress data that has the same item, so that it can reduce repeated database scans in the mining process and therefore this algorithm can run faster than the a priori algorithm [23].

This research will analyze the pattern of violence using the FP-Growth algorithm. Based on this literature study, this research contributes to a specific field, namely helping the DP3AKB to analyze patterns of violence which can later be used as solutions and also recommendations in dealing with problems of violence properly and on target.

## 2. RESEARCH METHOD

The research discussed is a study to build a website that will analyze patterns of violence using the association rule method, the Fp-Growth algorithm.

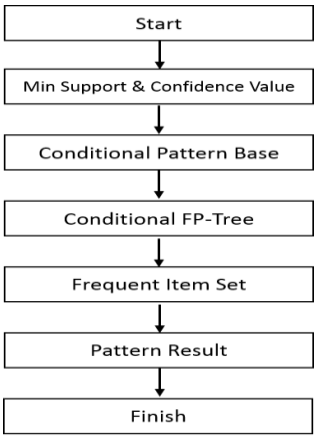


Figure 1. How the Fp-Growth Algorithm Works.

### *2.1. Association Rule Techniques*

Association analysis or association rule mining is a data mining technique for finding associative rules between combinations of items. In determining an association rule, there is a measure that states that any information or knowledge is considered interesting (interestingness measure). This measure is obtained from the results of data processing certain calculations. To measure the interestingness measure, the following variables can be used:

a. Support

$$\text{Support (A)} = \frac{\text{Number of Transactions Containing A}}{\text{Total Transactions}} \quad (1)$$

b. Confidence

$$\text{Confidence } (A \rightarrow B) = \frac{\text{Number of Transactions Containing A \& B}}{\text{Number of Transactions Containing A}} \quad (2)$$

### *2.2. FP-Growth Algorithm*

The FP-Growth algorithm is a development of the Apriori algorithm. So the shortcomings of the Apriori algorithm are corrected by the FP-Growth algorithm. Frequent Pattern Growth (FP-Growth) is one alternative algorithm that can be used to determine the set of data that most often appears (frequent itemset) in a data set. Apriori algorithm requires generating candidates to get frequent itemsets. However, the FP-Growth algorithm generates candidates not done because FP-Growth uses the concept of tree development in the search for frequent itemsets. That is what causes the FP-Growth algorithm to be faster than the Apriori algorithm. The characteristic of the FP-Growth algorithm is the data structure used is a tree called FP-Tree. By using FP-Tree, the FPgrowth algorithm can directly extract frequent Itemset from FP-Tree. The FP-Growth method can be divided into 3 main stages, as follows:

1. The generation phase of a conditional pattern base,
2. The generation phase of FP-Tree conditionals, and
3. The frequent itemset search stage.

### *2.3. Lift Ratio*

Lift ratio is used to measure how important the rule has been formed based on the value of support and confidence. Lift ratio is a comparison between confidence and the value of benchmark confidence. Benchmark confidence is the comparison between the sums of all items consequent to the total number of transactions.

$$\textit{Benchmark Confidence} = \frac{NC}{N} \quad (3)$$

$$\textit{Lift Ratio} = \frac{\textit{Confidence } (A, C)}{\textit{Benchmark confidence } (A, C)} \quad (4)$$

Where NC is the number of transactions with items of consequence, and N is the number of database transactions. If the lift ratio value is greater than 1, it shows the benefits of the rule. The higher the lift ratio, the greater the association strength.

### *2.4. Data Processing*

In this study, the source of the data used is the forms of violence in 2015, 2016 and 2017. The data will be processed to produce knowledge that can be used to find out what forms of violence often occur. Data obtained in the form of Excel, the following examples of data.

| | Kota/Kabupaten | Fisik | Psikis | Seksual | Eksploitasi | Trafficking | Lainnya | Penelantaran | KDRT |
|---|---|---|---|---|---|---|---|---|---|
| 2 | Kabupaten Banjarnegara | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| 3 | Kabupaten Banyumas | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 4 | Kabupaten Batang | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 5 | Kabupaten Blora | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 6 | Kabupaten Boyolali | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 |
| 7 | Kabupaten Brebes | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | Kabupaten Cilacap | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 9 | Kabupaten Demak | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 10 | Kabupaten Grobogan | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 11 | Kabupaten Jepara | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
| 12 | Kabupaten Karanganyar | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 13 | Kabupaten Kebumen | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 14 | Kabupaten Kendal | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 |
| 15 | Kabupaten Klaten | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 16 | Kabupaten Kudus | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 17 | Kabupaten Magelang | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 |
| 18 | Kabupaten Pati | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 |
| 19 | Kabupaten Pekalongan | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 20 | Kabupaten Pemalang | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 21 | Kabupaten Purbalingga | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 22 | Kabupaten Purworejo | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |
| 23 | Kabupaten Rembang | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 24 | Kabupaten Semarang | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 |

Figure 2. Data Forms of Violence

In the picture above is an example of violence data that has physical, psychological, and sexual, exploitation, trafficking, other variables, neglect, and domestic violence.

## 3. RESULTS AND DISCUSSION

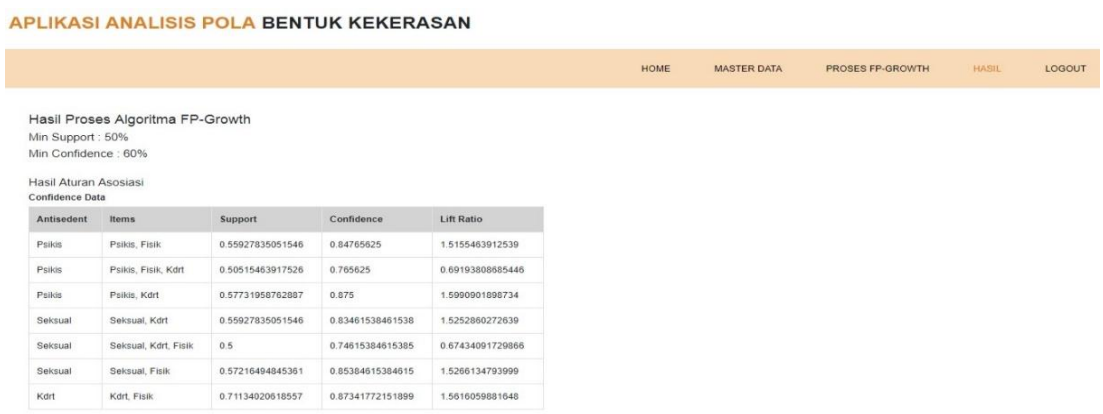


Figure 3. Admin Display Result

In Figure 3 is the result of the admin display that provides rule information, support values, confidence values, and lift ratio.


Figure 4. User Display Result

Figure 4 is the result of the user display that provides rule information and recommendations (solution).

In the violence data amounted to 420 data with forms of violence, namely physical, psychological, sexual abuse, exploitation, trafficking, neglect, domestic violence, and others that have been processed in the application of the analysis of patterns of violence using the FP-growth algorithm, the best rule results as shown on table 1 with min values support 50% and 60% min confidence, with the following rule:

Table 1. The Best Rule Result

| Rule | Support | Confidence | Lift Ratio |
|---|---|---|---|
| If there is psychological violence, then physical violence | 0,559 | 0,827 | 1,515 |
| If there is psychological violence, then also physical violence & household violence | 0,765 | 0,765 | 0,691 |
| If there is psychological violence, then also household violence | 0,577 | 0,875 | 1,599 |
| If there is sexual violence, then also household violence | 0,559 | 0,834 | 1,525 |
| If there is sexual violence, then also household violence & physical violence | 0,5 | 0,746 | 0,674 |
| If there is sexual violence, then also household violence | 0,572 | 0,853 | 1,526 |
| If there household violence, then also physical violence | 0,711 | 0,873 | 1,561 |

This results in the 7 best rules and produces a lift ratio value. If the value of the lift ratio is greater than 1, it indicates the benefits of the rule. It also says that the rule is good. In the rule above, there are 5 out of 7 rules whose ratio values are greater than one, which means that of the best rules above there are 71% good rules.

In each rule also produces recommendations such as the table 2:

Table 2. The Recommendation or Solution Results

| Rule | Solution |
|---|---|
| If there is psychological violence, then physical violence | The socialization on "self-love" or about self-education and about character education then facilitates psychologists to provide counseling on psychology, then counseling about procedures for reporting acts of violence to the authorities / related, making online sites for reporting violence. |
| If there is psychological violence, then also physical violence & household violence | Socialization about "self-love" or about self-respect and about character education then facilitates psychologists to provide counseling, socialization about the importance of communication between families, then counseling about procedures for reporting acts of violence to the authorities / related, making online sites for reporting violence. |
| If there is psychological violence, then also household violence | Socialization about "self -love" or about self-respect and about character education then facilitates psychologists to provide counseling, socialization about the importance of communication between families, then counseling about procedures for reporting acts of violence to the authorities / related, making online sites for reporting violence. |
| If there is sexual violence, then also household violence | Socialization about sex education that must be started early, and socialization about the importance of communication between family members, create an online site for reporting violence. |
| If there is sexual violence, then also household violence & physical violence | Socialization about sex education that must be started early, socialization about the importance of communication between family members, and counseling on procedures for reporting acts of violence to authorities / related parties, making online sites for reporting violence. |
| If there is sexual violence, then also household violence | Dissemination of sex education that must start early, then counseling about procedures for reporting acts of violence to the authorities / related, creating an online site for reporting violence. |
| If there household violence, then also physical violence | Socialization about the importance of communication between families, then counseling about procedures for reporting acts of violence to the authorities / related, making online sites for reporting violence. |

In each rule, recommendations are given in the hope that the parties responsible or involved in matters of violence can make the best possible and rightful prevention efforts.

## 4. CONCLUSION

Based on the discussion and data processing of violence in Central Java Province, it can be concluded that: the application of the violence pattern analysis with the FP-Growth algorithm that is built can run well and in accordance with the expected objectives.

Through data on violence analysis in Central Java Province, amounting to 420 in the form of violence, namely physical, psychological, sexual violence, exploitation, trafficking,

neglect, domestic violence, and others, this data from 2015, 2016 and 2017 obtained the best 7 rules at the value level, minimum support = 50% and minimum confidence value = 60%. From the results of 7 rules, there are 5 rules that are categorized as good because the lift ratio value is greater than 1. If the lift ratio value is greater than 1, it means that the resulting pattern is getting stronger.

The application of violence pattern analysis using the FP-Growth algorithm, it can help DP3AKB to find any forms of violence that often occur, so that they can take appropriate actions to deal with and prevent violence. In this application, there are also recommendations on each rule, these recommendations can help officers to prevent violence properly and on target.

For further research, the results of testing can be further developed by increasing the amount of data. More data are processed so as to produce better association rules and generate new knowledge in violence prevention. The results of recommendations or Solutions from the application can be more numerous and varied. The application process in the data upload section can be directly from the DP3AKB server or related agencies so that data can be accessed in real time and up to date. Admin is expected to be able to continue to maintain the application regularly. Besides that, you can also try to use other data mining algorithms as a comparison, so that more efficient and effective algorithms can be found.

## *REFERENCES*

[1] T. Santoso, *Kekuasaan Dan Kekerasan*, vol. 14, no. 4. 2001.

[2] F. Gorunescu, *Data Mining - Concepts, Models and Techniques*, vol. 12. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.

[3] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining - Practical Machine Learning Tools and Techniques (Fourth Edition)*, Fourth Edi., vol. 53, no. 9. Morgan Koufmann & Elsevier, 2017.

[4] M. Bramer, *Principles of data mining fourth edition*, vol. 30, no. 7. 2020.

[5] A. Luthfiarta, J. Zeniarja, E. Faisal, and W. Wicaksono, "Prediction on Deposit Subscription of Customer based on Bank Telemarketing using Decision Tree with Entropy Comparison," *J. Appl. Intell. Syst.*, vol. 4, no. 2, pp. 57–66, 2019.

[6] J. Zeniarja, K. Widia, and R. R. Sani, "Penerapan Algoritma Naive Bayes dan Forward Selection dalam Pengklasifikasian Status Gizi Stunting pada Puskesmas Pandanaran Semarang," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 1–9, 2020.

[7] A. Luthfiarta *et al.*, "Classification of Governer's Public Report from SMS LaporGub Using Naive Bayes Classifier Method," *Proc. - 2018 Int. Semin. Appl. Technol. Inf. Commun. Creat. Technol. Hum. Life, iSemantic 2018*, pp. 214–218, 2018.

[8] I. Purnamasari, F. Handayanna, E. Arisawati, L. S. Dewi, E. G. Sihombing, and Rinawati, "The Determination Analysis of Telecommunications Customers Potential Cross-Selling with Classification Naive Bayes and C4.5," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020.

[9] B. Changkakati and C. Das, "Data mining techniques in hr analytics: A review of domain specific concepts and technicalities," *Int. J. Sci. Technol. Res.*, vol. 9, no. 3, pp. 4358–4362, 2020.

[10] O. Stit, J. Riffi, A. Yahyaouy, and H. Tairi, "Comparative Study of Different Association Rule Methods," in *Colloquium in Information Science and Technology, CIST*, 2018, vol. 2018-Octob, pp. 323–327.

[11] W. N. Ismail, M. M. Hassan, and H. A. Alsalamah, "Context-Enriched Regular Human Behavioral Pattern Detection from Body Sensors Data," *IEEE Access*, vol. 7, no. 1, pp. 33834–33850, 2019.

[12] T. P. Hong, C. Y. Lin, W. M. Huang, K. S. M. Li, L. S. L. Wang, and J. C. W. Lin, "Using Tree Structure to Mine High Temporal Fuzzy Utility Itemsets," *IEEE Access*, vol. 8, pp. 153692–153706, 2020.

[13] P. Naresh and R. Suguna, "Association rule mining algorithms on large and small datasets: A comparative study," in *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, 2019, pp. 587–592.

[14] D. Winarti *et al.*, "Analisis Data Mining Dengan Algoritma Fp-Growth Dalam Mendukung Strategi Promosi," *Simtika*, vol. 1, no. 1, pp. 27–31, 2018.

[15] G. Gunadi and D. I. Sensuse, "Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth ( Fp-Growth ) :," *Telematika*, vol. 4, no. 1, pp. 118–132, 2012.

[16] E. Faisal, J. Zeniarja, and D. A. N. Sa'adah, "Pola Beli Konsumen menggunakan Algoritma FP-Growth untuk Rekomendasi Promosi penjualan pada Batik Nadya Pekalongan," in *SeNTIK - STMIK JAKARTA STI&K*, 2017.

[17] M. Hossain, A. H. M. S. Sattar, and M. K. Paul, "Market basket analysis using apriori and FP growth algorithm," in *2019 22nd International Conference on Computer and Information Technology, ICCIT 2019*, 2019.

[18] Islamiyah, P. L. Ginting, N. Dengen, and M. Taruk, "Comparison of Priori and FP-Growth Algorithms in Determining Association Rules," in *ICEEIE 2019 - International Conference on Electrical, Electronics and Information Engineering: Emerging Innovative Technology for Sustainable Future*, 2019, pp. 320–323.

[19] M. S. Kaysar, M. A. Bin Khaled, M. Hasan, and M. I. Khan, "Word Sense Disambiguation of Bengali Words using FP-Growth Algorithm," in *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019*, 2019, pp. 1–5.

[20] R. Fitria, W. Nengsih, and D. H. Qudsi, "Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas," *J. Sist. Inf.*, vol. 13, no. 2, p. 118, 2017.

[21] I. Astrina, M. Z. Arifin, and U. Pujianto, "Penerapan Algoritma FP-Growth dalam Penentuan Pola Pembelian Konsumen pada Kain Tenun Medali Mas," *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 1, pp. 32–40, 2019.

[22] Fadlina, "Data Mining Untuk Analisa Tingkat Kejahatan Jalanan Dengan Algoritma Association Rule Metode Apriori," *Inf. dan Teknol. Ilm.*, vol. 3, no. 1, pp. 144–154, 2014.

[23] D. Y. Hardiyanti, H. Novianti, and A. Rifai, "Penerapan Algoritma Fp-Growth Pada Sistem Informasi Perpustakaan," *J. Comput. Eng. Syst. Sci.*, vol. 3, no. 1, pp. 75–77, 2018.