Implementation Data Mining For Level Analysis Traffic Violation By Algorithm Association Rule

Oris Krianto Sulaiman[1], Indah Purnama Sari[2], Andy Satria[3]
[1]Department of Informatics Engineering, Universitas Islam Sumatera Utara, Indonesia
[2]Department of Information Technology, Universitas Muhammadiyah Sumatera Utara, Indonesia
[3]Department of Information Technology, Universitas Dharmawangsa, Indonesia

ABSTRACT

Police with regard to its function as a protector of society is expected able to take action in responding to traffic violations by conducting raid operations in uncovering violations that can be committed using analysis technique of habits how often traffic violations are committed. Detection Traffic violations that often occur simultaneously are called Association Rules.

Keyword : Data mining, FP-Growth

Corresponding Author:
Oris Krianto Sulaiman,
Department of Informatics Engineering,
Universitas Islam Sumatera Utara,
Jalan Sisingamangaraja No 59 Medan, Indonesia.
Email: oris.ks@ft.uisu.ac.id

1. INTRODUCTION

Traffic violation already entrenched in society, is no exception occurred in the city of Padang. With its status as provincial capital and city large which has a density level the highest population in Sumatra Province West, of course, is the city of Padang as a city that has levels highest density of motorized vehicles when compared to other regions. Padang city is one of the cities which has data on the number of accidents The largest is proven from data from 2012 to 2012 2013, the city of Padang is in a ranking position first most accident cases in West Sumatra Province. In 2012 , in the city of Padang recorded 540 cases accidents, with details of 80 victims passed away [1].

Police related function as a protector of society expected to take action in dealing with traffic violations by carrying out deep raid operations uncover possible violations done using technique analysis of how often habits traffic violation is committed. Traffic violation detection often occur simultaneously Association Rule.

Association analysis is also known as a data mining technique form the basis of various data techniques other mining. Especially one of the stages of association analysis is called analysis high frequency patterns mining) has attracted the attention of many researchers to produce efficient algorithms. The importance of an associative rule can be determined by two parameters, support (support value), namely the percentage the combination of those items in the database and confidence [2].

In this study, the algorithm that used is Frequent Pattern-Growth (FP-Growth) namely the development of A priori method which is one alternatives to determine the data set the most frequently occurring (frequent itemset) in a data set with generate a Tree data structure or called the Frequent Pattern Tree (FPTree) [3].

2. RESEARCH METHOD/MATERIAL AND METHOD/LETERATURE REVIEW (10 PT)

A. Pattren of FP-Growth

Frequent Pattern Growth (FP-Growth) is differentapproach for finding the most frequently occurring data set (frequent itemset) in a data set. The apriori algorithm was developed into the FP-Growrth algorithm. As a result, the FP-Growth algorithm compensates for the limitations of the Apriori method.

In order to find common itemset, FP-Growth employs the concept of tree construction. The FP-Growth algorithm is distinguished by the usage of a data structure known as the FP-tree. FP-Growth algorithm may directly extract frequent itemset from FP-Tree using FP-Tree. The FP-tree is a data storage structure that is compressed. Because there may be transaction with the same item in each mapped transaction, the pathways allow for overlap. The compression process with the FP-tree data structure will be more effective the more transaction data that has the same item. FP-tree has the advantage of requiring just two scans of transaction data, which has shown to be quite efficient. Let i= {a1, a2, ..., an} is a collection of items. And transaction database DB = {T1, T2, ..., Tn}, where Ti (i € [1...n]) is a set of transactions containing items in I. While support is a counter for the number of times a pattern appears in a transaction. With the use of an FP-tree structure, FP-Growth attempts to tackle the challenge of discovering a common pattern with a minimal threshold of support count £.

B. Association Rule

Association analysis or association rule mining is a data mining technique for find the associative rule between a combination of items. Association rules algorithm will use training data, accordingly with the notion of data mining, for generate knowledge. Knowledge what to produce in the rules association? Knowledge to know frequently purchased shopping items together at a time. Rule association "if...then"
resulting from the association rule function [5,6].

3. RESULTS AND DISCUSSION

A. Data mining analysis with FP-Growth Algorithm

In this study apply Data Mining association rules with FPGrowth algorithm in analyzing data on traffic violations at Polresta Field to get the pattern violation that occurred. Existing pattern describe cause and effect traffic violations at Polresta.

In the rule search process association of all traffic violations Existing itemsets are considered as input on Data Mining. All itemsets are processed using the FP-Growth algorithm thus forming an association relationship between itemset by meeting the minimum support and minimum confidence to set valid data. Generated data analyzed by association rule "if….then". To reach all stages that's the first step to do is collect data. To perform data mining traffic violations, author using the FP-Growth algorithm. For the frequency of occurrence of each item from transaction data can be seen in the table following :

Table 1. Table of Transaction Data Frequency

Tid	Itemset	Frequency
1	Student	6
2	Employee	5
3	Driver	3
4	Car	4
5	Motorcycle	7
6	Truck Buses	3
7	SIM	ii
8	STNK	6
9	Equipment	10
10	Signs	6

B. Formation of FP-Tree

TID	Dataset Pelanggaran
1	(a.e^Td)
2	CMj)
3	
4	
5	(a.e^TU)
6	(a,d^4i4j)
7	<t=esl
S	
9	(=£g£)
10	(M4)
11	(a,e^Ji4)
12	
13	
14	

From the table above, the next steps are have to do is form the FPtree path. Where is the result of the formation of line 1 up to 14 can be seen in the following figure:

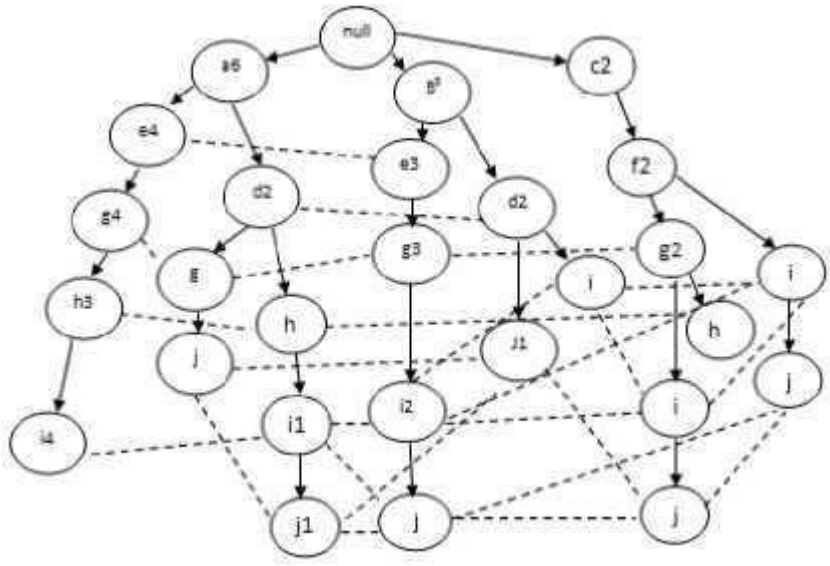

Figure 1. Decission Tree FP-Growth

C. Application of FP-Growth

After the FP-tree building stage data on existing offenders, next applied the FP-growth algorithm to find frequent itemsets. FPgrowth Algorithm divided into three steps viz Conditional Pattern Base, Conditional FPtree, and Frequent Itemsets.

Conditional Pattern Base is subdatabase that contains the path prefix (path prefix) and a suffix pattern (suffix pattern). Generation of conditional pattern base obtained through the FP-tree that has been built. From the results of the formation of FP-Tree before then the steps taken is to generate conditional FP-Tree for paths containing the suffix j up to the suffix a.

Result of FP-Tree generation then we get the Conditional Pattern Base. The results can be seen in the following table:

Table 2. Conditional Pattern Base

Suffix	Conditional Pattern Base
j	{(i,d,g:l),(i,dJixl),(b,e,ga:l),(b,d:l),M&
i	{(a,e:gth:4),(a,dji: 1),(b,e,g:2),(b,d: 1
h	{(a,C,s:3),(a,d:l),(c,f,g:l)>
S	{(a,e:4),(a,d:l),(b,e:3),(c,f:2)}
f	{(c:2)}
E	{(a:4),(b:3)}
D	K*:2Hb:2)}

Support count of each item on conditional pattern base is added up, then items that have a support count greater equal to the minimum support count % will be raised with conditional FP-tree.

Suffix	Conditional Pattern Base	Conditional FP-
J		
I	{(a,egJi:4),(a,d3i:l),(b,^:2),(b,	{(a:5,b:3,
R	!(w3i(a,±l),(ci5l)}	{(a:2,s:4)}
G	{(a ,e:4),(a ,d: 1	{(a:2,e:7,f
F	tfc:2)3	
E	{MW>3»	{(a:4,b:3)
D		{(a:2,b:2)

If Conditional FP-tree is a single path, then get the frequent itemset with do a combination of items for each conditional FP-tree. If not track single, then FPgrowth is generated recursively.

Suffix	*Conditional Pattern Base*	*Frequent Itemset*
J	{(Mg lKxdtoil Id xgjim.d: 1	(dj):3,(gj>
I	((3,egdi:4Mxd4i:l),(b,eg:2),	(a4):5,(bj)
H		C&h)4
G	{(a,e:4),(a,d:l).{b,e:3),(c,f:2	(<&?
F	{(<=2)5	
E		(a,e):4,
D	{($_a$:2),(b:2)J	

Then calculate the support value and the confidence. Value calculated items the support and confidence are: {d,j}, {g,j}, {a,i}, {b,i}, {g,i}, {h,i}, {g,h}, {e,g}, {a,e}, {b,e}.

$$\text{Support}(A) = \frac{\text{Jumlah Transaksi mengandung A}}{\text{Total Transaksi}}$$

sedangkan nilai support dari 2 item diperoleh dari rumus berikut:

$$\text{Support}(A \cap B) = \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Total Transaksi}}$$

Calculating Confidence value :

$$\text{Confidence} - P(B \mid A) - \frac{\text{Jumlah Transaksi mengandung A dan B}}{\text{Jumlah Transaksi mengandung A}} \quad \text{III}$$

The results of the value of support and confidence value for more details can seen in the following table:

Jika Melanggar	Maka Akan Melanggar	Support	Confidence
Mobil	Rambu-rambu	21,43%	75%
Rambu-rambu	Mobil	21,43%	50%
SIM	Rambu-rambu	21,43%	27,27%
Rambu-rambu	SIM	21,43%	50%
Pelajar	Perlengkapan	35,71%	83,33%
Perlengkapan	Pelajar	35,71%	50%
Mobil	Perlengkapan	21,43%	60%
Perlengkapan	Mobil	21,43%	30%
SIM	Perlengkapan	42,86%	54,55%
Perlengkapan	SIM	42,86%	60%
STNK	Perlengkapan	35,71%	83,33%
Perlengkapan	STNK	35,71%	50%
SIM	STNK	28,57%	36,36%
STNK	SIM	28,57%	66,67%
Motor	SIM	50%	100%
SIM	Motor	50%	63,64%
Pelajar	Motor	28,57%	66,67%
Motor	Pelajar	28,57%	57,14%
Karyawan	Motor	21,43%	60%
Motor	Karyawan	21,43%	42,86%

Rule can be seen in the following table :

IjJbt.Xtelanggar	\tMsa Akan Melanggar	Support	Confidence
Pelajar	Parian Ekap	35,71%	83,33%
SIM	Parlanskapan	42,86%	54,55%
STNK	Parian Ekap	35,71%	83,33%
Motor	SIM	50%	100%

From the rule generated above, if item a (student) then violates item I (equipment), if item g (SIM) then violating item i (Equipment), if item h (STNK) then violates the item (Equipment), if item e (Motor) then breaking item g (SIM).

4. CONCLUSION

The FP-Growth algorithm is wrong one correct method to apply on traffic violation data due assist the police in analysis pattern of violations that are frequently violated. Matter this helps the police force to improve police performance. The results of the analysis show that the most frequent violations according to the data obtained. Matter This happens because it is still lacking concern for motorists to comply existing rules.

REFERENCES

[1] Sari, I.P., Al-Khowarizmi, A., & Batubara, I.H (2021). Optimization of the FP-Growth Algorithm in Data Mining Techniques to Get the Electric Power Theft Pattern for the Development of Smart City. *4th International Conference of Computer and Informatics Engineering (IC2IE)*, 293-298.

[2] Pramudiono, Iko. 2003. Pengantar Data Mining : Menambang Permata Pengetahuan di Gudang Data. http:// www.ilmukomputer.com, tanggal akses 24 April 2014.

[3] Sari, I.P., Al-Khowarizmi, A., & Batubara, I.H (2021). Cluster Analysis Using K-Means Algorithm and Fuzzy C-Means Clustering For Grouping Students' Abilities In Online Learning Process. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, 2(1), 139-144.

[4] Erwin, 2009. Analisis Market Basket Dengan Algoritma Apriori dan FPGrowth. *Jurnal Generic*, vol. 4.

[5] Sari, I.P., Batubara, I.H., & Al-Khowarizmi, A (2021). Sensitivity Of Obtaining Errors In The Combination Of Fuzzy And Neural Networks For Conducting Student Assessment On E-Learning. *International Journal of Economic, Technology and Social Sciences (Injects)*, 2(1), 331-338.

[6] Sari, I.P., Fahroza, M.F., Mufit, M.I., & Qathrunad, I.F (2021). Implementation of Dijkstra's Algorithm to Determine the Shortest Route in a City. *Journal of Computer Science, Information Technology and Telecommunication Engineering,* 2(1), 134-138.

[7] Batubara, I.H., Saragih, S., Syahputra, E., Armanto, D., Sari, I.P., Lubis,B.S., & Siregar, E.F.S (2022). Mapping Research Developments on Mathematics Communication: Bibliometric Study by VosViewer. AL-ISHLAH: *Jurnal Pendidikan* 14(3), 2637-2648.

[8] Sari, I.P., Al-Khowarizmi, A.K., & Batubara, I.H. (2021). Analisa Sistem Kendali Pemanfaatan Raspberry Pi sebagai Server Web untuk Pengontrol Arus Listrik Jarak Jauh. InfoTekJar: *Jurnal Nasional Informatika dan Teknologi Jaringan,* 6 (1), 99-103.

[9] Hariani.,P.P, Sari.,I.P, & Batubara., I.H. (2021). Implementasi e-Financial Report BUMDes. *IHSAN: JURNAL PENGABDIAN MASYARAKAT*, 3 (2), 169-177.

[10] Sari, I.P., Basri, Mhd., Ramadhani, F., & Manurung, A.A. (2023). Penerapan Palang Pintu Otomatis Jarak Jauh Berbasis RFID di Perumahan. *Blend Sains Jurnal Teknik*, 2(1), 16-25.

[11] Batubara, I.H., & Sari, I.P. (2021). Penggunaan software geogebra untuk meningkatkan kemampuan pemecahan masalah matematis mahasiswa. *Scenario (Seminar of Social Sciences Engineering and Humaniora)*, 398-406

[12] Sari, I.P., & Batubara, I.H. (2020). Aplikasi Berbasis Teknologi Raspberry Pi Dalam Manajemen Kehadiran Siswa Berbasis Pengenalan Wajah. *JMP-DMT* 1(4), 6.

[13] Sari, I.P., Al-Khowarizmi, A.K., Ramadhani, F., & Sulaiman, O.K. (2023). Implementation of the Selection Sort Algorithm to Sort Data in PHP Programming Language. *Journal of Computer Science, Information Technology and Telecommunication Engineering*, 4(1).

[14] Batubara, I.H., Sari,I.P., Hariani, P.P., Saragih, M., Novita, A., Lubis, B.S., & Siregar, E.F.S. (2021). Pelatihan Software Geogebra untuk Meningkatkan Kualitas Pembelajaran Matematika SMP Free Methodist 2. *Martabe: Jurnal Pengabdian Kepada Masyarakat,* 4(3), 854-859.

[15] Sari., I.P, Batubara., I.P, Al-Khowarizmi., A, & PP Hariani. (2022). Perancangan Sistem Informasi Pengelolaan Arsip Digital Berbasis Web untuk Mengatur Sistem Kearsipan di SMK Tri Karya. *Wahana Jurnal Pengabdian kepada Masyarakat* 1 (1), 18-24.

[16] Batubara., I.H, Sari., I.P, EFS Siregar, & BS Lubis. (2021). Meningkatkan Kemampuan Penalaran Matematika Melalui Metode Penemuan Terpandu Berbantuan Software Autograph. *Seminar Nasional Teknologi Edukasi Sosial dan Humaniora* 1 (1), 699-705.

[17] Sari., I.P, A Syahputra, N Zaky, RU Sibuea, & Z Zakhir. (2022). Perancangan sistem aplikasi penjualan dan layanan jasa laundry sepatu berbasis website. *Blend sains jurnal teknik* 1 (1), 31-37.

[18] Sari., I.P, A Azzahrah, FQ Isnaini, L Nurkumala, & A Thamita. (2022). Perancangan sistem absensi pegawai kantoran secara online pada website berbasis HTML dan CSS. *Blend sains jurnal teknik* 1 (1), 8-15.

[19] Ramadhani., F, & Sari., I.P. (2021). Pemanfaatan Aplikasi Online dalam Digitalisasi Pasar Tradisional di Medan. *Prosiding Seminar Nasional Kewirausahaan* 2 (1), 806-811.

[20] Sari.,I.P, & Ramadhani., F. (2021). Pengaruh Teknologi Informasi Terhadap Kewirausahaan Pada Aplikasi Perancangan Jual Beli Jamu Berbasis WEB. *Prosiding Seminar Nasional Kewirausahaan* 2 (1), 874-878.

[21] Sari., I.P, A Jannah, AM Meuraxa, A Syahfitri, & R Omar. (2022). Perancangan Sistem Informasi Penginputan Database Mahasiswa Berbasis Web. *Hello World Jurnal Ilmu Komputer* 1 (2), 106-110.

[22] Hutasuhut, B.K., Sari, I.P., & Al-Khowarizmi, A (2023). Analysis the Effect of Digitalization and Technology on Web-Based Entrepreneurship. *Journal of Computer Science, Information Technology and Telecommunication Engineering* 4(1).

[23] Sari., I.P, & Batubara., I.H. (2021). Perancangan Sistem Informasi Laporan Keuangan Pada Apotek Menggunakan Algoritma K-NN. *Seminar Nasional Teknologi Edukasi dan Humaniora (SiNTESa)* 1 (2021 - ke 1.

[24] Ramadhani., F, A Satria, & Sari., I.P. (2022). Aplikasi Internet Berbasis Website sebagai E-Commerce Penjualan Komponen Sport Car. *Blend Sains Jurnal Teknik* 1 (2), 69-75.

[25] Sari., I.P, & Batubara., I.H. (2021). User Interface Information System for Using Account Services (Joint Account) WEB-Based. *International Journal of Economic, Technology and Social Sciences (Injects),* 462-469.

[26] PP Hariani, Sari., I.P, & Batubara., I.H. (2021). Android-Based Financial Statement Presentation Model. *JURNAL TARBIYAH* 28 (2), 1-16.

[27] Sari., I.P, Batubara., I.H, & M Basri. (2022). Implementasi Internet of Things Berbasis Website dalam Pemesanan Jasa Rumah Service Teknisi Komputer dan Jaringan Komputer. *Blend Sains Jurnal Teknik* 1 (2), 157-163.

[28] Ramadhani, F., Satria, A., & Sari, I.P (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer* 2(2), 58-62.

[29] Sari., I.P, Al-Khowarizmi., A, & Batubara., I.H. (2021). Implementasi Aplikasi Mobile Learning Sistem Manajemen Soal dan Ujian Berbasis Web Pada Platform Android. *IHSAN: JURNAL PENGABDIAN MASYARAKAT* 3 (2), 178-183.

[30] Batubara, I.H., Saragih, S., Simamora, E., Napitupulu, E.E., Sari, I.P. (2022). Analysis of Student's Mathematical Communication Skills through Problem Based Learning Models Assisted by Augmented Reality. Budapest *International Research and Critics Institute-Journal (BIRCI-Journal)*, 5(1), 1024-1037.