

## ANALISIS POLA KECELAKAAN LALU LINTAS MENGGUNAKAN ALGORITMA K-MEANS DAN FP-GROWTH STUDI KASUS: POLRES SOLOK

<sup>1</sup>M. Afdal, <sup>2</sup>Risky Prana Disastra

<sup>1,2</sup>Program Studi Sistem Informasi, Fakultas Sains dan Teknologi UIN Suska Riau

Jl. HR Soebrantas KM.18 Panam Pekanbaru - Riau

Email: <sup>1</sup>m.afdal@uin-suska.ac.id, <sup>2</sup>11553100437@students.uin-suska.ac.id

### ABSTRAK

Kabupaten Solok merupakan salah satu kabupaten yang ada di Sumatera Barat yang memiliki jumlah kecelakaan yang terbilang tinggi, yakni sebanyak 497 kejadian selama tahun 2015-2018, hal ini disebabkan oleh jumlah kendaraan yang kian meningkat, perkembangan transportasi sangat mempengaruhi resiko akan terjadinya kecelakaan dan memerlukan penanganan serius karena dapat menyebabkan kerugian yang besar. Pada unit kecelakaan (Laka) Polres Solok, data ini digunakan untuk pelaporan sehingga bisa diketahui berapa banyak korban setiap tahunnya. Karena banyaknya data kecelakaan maka diperlukan pengelompokan data dengan metode clustering. Proses ini bertujuan untuk mengelompokkan berdasarkan karakteristik tiap objek, data yang sudah dikelompokkan akan diolah dan diproses sehingga dapat memberikan sebuah rekomendasi dalam mengurangi resiko terjadinya kecelakaan. Penelitian ini menggunakan metode algoritma K-Means untuk clustering dan FP-Growth untuk mendapatkan aturan asosiasi. Dari data yang diproses maka didapatkan hasil berupa kecelakaan terjadi pada usia muda (16-35 Tahun), keadaan jalan sepi, jenis kendaraan sepeda motor, pengendara tidak memiliki sim, pengendara berjenis kelamin laki-laki, jenis luka dengan kondisi luka ringan. Dalam menanggulangi kecelakaan lalu lintas dapat dilakukan pada usia muda dengan memberikan arahan dan sosialisasi pada sekolah serta unit lainnya, sedangkan untuk keadaan jalan sepi yaitu memberikan fasilitas umum berupa penerangan lampu pada jalan dan infrastruktur yang terdekat.

**Kata Kunci:** *Association Rule, Data Mining, K-Means, FP-Growth, Laka lintas*

### A. PENDAHULUAN

Kecelakaan lalu lintas seringkali terjadi karna kelalaian manusia, seperti bermain smartphone ketika berkendara, berkendara saat keadaan mengantuk, dan tak mengamati rambu lalu lintas. Sebagian tahun terakhir jumlah kecelakaan alami kenaikan, sehingga memerlukan perhatian yang dikhususkan dan penanggulangan yang sungguh-sungguh untuk mengurangi efek kecelakaan [14].

Pemerintah Kabupaten Solok harus lebih memperhatikan kecelakaan lalu lintas yang terjadi belakangan ini, dimana kecelakaan lalu lintas merupakan masalah yang memerlukan penanganan serius yang dapat menyebabkan kerugian yang besar. Pada tahun 2015-2018 mencatat terjadi 497 kasus kecelakaan yang menimbulkan 509 orang mengalami luka ringan, 377 orang mengalami luka berat dan 124 orang meninggal dunia. Angka yang terdapat pada data kecelakaan ini hanyalah yang melaporkan pada pihak berwenang saja, karna pada kenyataannya banyak masyarakat yang enggan melaporkan kejadian ini pada lembaga kepolisian setempat.

Jumlah korban kecelakaan cenderung meningkat setiap tahun. Peningkatan produksi kendaraan bermotor, terutama di Indonesia tidaklah demikian berkontribusi langsung terhadap bertambahnya nilai kasus kecelakaan lalu lintas. Tidak hanya dari meningkatnya jumlah kendaraan, aspek lain serupa kelalaian serta minimnya

pengetahuan ialah aspek utama pemicu terbentuknya kecelakaan lalu lintas. Bertambahnya kasus kecelakaan lalu lintas dari tahun ke tahun membutuhkan kepedulian serta tindakan penghindaran yang serius dimana kasus kecelakaan lalu lintas aspek kematian yang lumayan tinggi [14]. Penanganan kecelakaan lalu lintas bisa dicoba dengan menerapkan langkah-langkah pendekatan seperti sosialisasi, pengawasan rambu lalu lintas. Untuk bisa menemukan penyelesaian kegiatan penanggulangan yang sesuai, dibutuhkan data dan informasi mengenai kecelakaan lalu lintas yang ada pada saat ini. Adapun metode dan cara yang efektif dalam analisis dan bisa diterapkan terhadap data kecelakaan lalu lintas yaitu menggunakan teknik data mining. Karena didalam data mining ada metode dan teknik untuk memperoleh informasi yang luas.

Data mining adalah suatu cara dalam mengekstrasi pola sebuah penggunaan data yang sangat besar. Kumpulan data mempunyai banyak potensial yang digunakan sebagai tujuan yang tepat dalam mengambil keputusan yang tepat yang dilakukan dengan cara menganalisa dan mencari informasi sebaik mungkin yang ada pada data yang dipakai [4]. Dalam penelitian ini yang akan dipakai dalam memperoleh informasi yang luas ialah dengan teknik association rule mining. *Association rule mining* yakni sebuah teknik yang digunakan dalam ditemukannya aturan asosiatif antara gabungan himpunan item. Metode ini dipakai untuk

menemukan korelasi atau asosiasi setiap item setiap dataset item, dengan menerapkan support (nilai pendukung) serta confidence (nilai kepastian) dari setiap item yang digunakan [6]. Metode ini bisa diterapkan di data yang jumlah datanya sangat banyak contohnya saja terhadap data jual beli sebuah produk [9]. Penerapan Algoritma terhadap FP-Growth terletak pada pemrosesan asosiasi dalam menentukan pola keterkaitan data kecelakaan lalu lintas supaya bisa mencari serta menemukan pola kecelakaan yang sering terjadi pada saat ini. Frequent Pattern Growth (FP-Growth) adalah algoritma alternatif yang bisa dipakai dalam mencari data himpunan yang sering sekali terlihat (frequent itemset).

Pada algoritma apriori sangatlah dibutuhkan generate candidate dalam mencari nilai frequent itemsets. Oleh sebab itu, pada algoritma FP-Growth memakai cara pembuatan tree dalam mencari sebuah itemset. Hal inilah yang mengakibatkan algoritma FP-Growth lebih bagus dalam menemukan aturan asosiasinya dari pada Apriori. Dalam menganalisa keterkaitan data pada sejumlah data pada aturan asosiasi cenderung mengabaikan itemset besar sehingga untuk data yang besar menjadi kurang akurat. Untuk mengatasi masalah tersebut, atribut yang ada dikluster terlebih dahulu menggunakan algoritma K-Means guna mencari sebuah atribut umum dalam sebuah kelompok serta dalam mencari pola asosiasi di setiap kelompok, supaya memudahkan pencarian kombinasi yang sering terjadi supaya nilai yang didapatkan lebih terjamin akurasi. Penelitian ini bertujuan untuk menemukan dan mencari faktor apa saja dalam terjadinya kecelakaan lalu lintas terhadap data kecelakaan lalu lintas yang terjadi pada Kabupaten Solok pada tahun 2015-2018. Algoritma K-Means dan FP-Growth telah berhasil dilakukan pada penelitian sebelumnya untuk menentukan hubungan antar item seperti: (1) pencarian aturan asosiasi pada data rekomendasi belajar berdasarkan segmentasi akademik siswa [11]; (2) menemukan aturan asosiatif pada data transaksi minimarket [8]; (3) menentukan aturan asosiasi pada data kecelakaan lalu lintas [16]; (4) untuk memperoleh informasi mengenai pengenalan pola daerah strategi pengenalan kampus [15]. (5) mencari aturan asosiasi untuk menentukan pola hubungan kecelakaan lalu lintas [2]

Berdasarkan penjelasan tersebut, maka penelitian ini mengangkat judul “Analisis Pola Kecelakaan Lalu Lintas Menggunakan Algoritma K-Means dan Fp-Growth Studi Kasus: Polres Solok”.

## B. LANDASAN TEORI

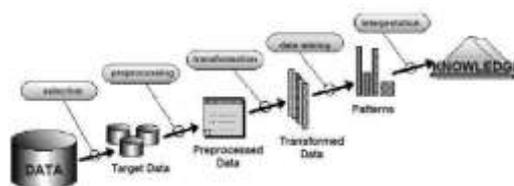
### B.1. Kecelakaan Lalu Lintas

Berdasarkan Undang-undang Nomor 22 tahun 2009 pasal 1 ayat 24 tentang lalu lintas dan angkutan jalan, Kecelakaan lalu lintas merupakan suatu peristiwa di jalan yang tidak terduga dan tidak

disengaja yang melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda. Faktor penyebab terjadinya kecelakaan lalu lintas yang sering diabaikan oleh pengguna jalan, yaitu: mengemudi saat keadaan mengantuk, menggunakan handphone saat berkendara, tidak mematuhi rambu-rambu lalu lintas.

### B.2. KDD (Knowledge Discovery in Database)

Data mining merupakan sebuah teknik dalam menggali informasi tersembunyi untuk memperoleh manfaat lebih dari data yang tersedia. Data mining dapat didefinisikan sebagai serangkaian proses untuk memperoleh pengetahuan. Salah satu tujuan dari data mining ialah mengklasifikasi, mengelompokkan, menemukan aturan asosiasi serta menemukan pola peramalan dimasa depan (predicting). Istilah data mining lebih dikenal dengan sebutan Knowledge Discovery from Database (KDD). Proses Knowledge Discovery in Database dapat dilihat pada Gambar 1 mulai dari pemilihan atribut data hingga terciptalah sebuah pengetahuan (knowledge). [4]



Gambar 1. Tahapan Proses KDD

### B.3. Data Mining

Data Mining atau dinamakan *knowledge discovery in database* (KDD) merupakan sesuatu kegiatan menemukan informasi dari sebuah data yang berdimensi sangat besar agar bisa menciptakan konsistensi pola serta hubungan pada sebuah itemset data, tahap pendataan informasi dapat dilakukan juga dengan penggunaan data histori sebelumnya [10]. Adapun 5 (lima) Tahap pengolahan pada proses Knowledge Discovery in Database (KDD) dapat dijabarkan dengan keterangan sebagai berikut: [4]

1. Data Selection
2. Data Preprocessing
3. Data Transformation
4. Data Mining
5. Interpretation/Evaluation

*Data mining* dapat dibagi menjadi beberapa kelompok antaralain:

1. Deskripsi
2. Estimasi
3. Prediksi
4. Klasifikasi
5. Pengklasteran (Clustering)
6. Asosiasi

### B.4. Association Rule

*Association rules* adalah teknik data mining untuk mendapatkan aturan asosiatif atau pola kombinasi dari sebuah barang. Misalkan dalam

aturan asosiatif dalam sebuah transaksi pembelian barang disuatu mini-market dapat dilihat seberapa besar kemungkinan konsumen membeli suatu barang bersamaan dengan barang lainnya, misalnya: saat membeli roti bersamaan dengan selai. Karena itu berasal dari tentang basis data transaksi pelanggan untuk menilai kebiasaan suatu produk yang dibeli secara bersamaan dengan produk lainnya, aturan asosiasi tersebut disebut dengan analisis keranjang pasar [12], dalam menentukan aturan asosiasi, ada ukuran yang menarik (ukuran kepercayaan) yang diperoleh hasil memproses data dengan menggunakan kalkulasi tertentu. Secara umum terdapat dua tipe ukuran, yaitu: [12]. Perhitungan nilai *support* dan *confidence*.

### B.5. Clustering

Clustering adalah proses pengelompokan data dalam sebuah kelas atau klaster sehingga pada sebuah data terhadap sebuah klaster mempunyai jenis keserupaan yang sangat tinggi antara data satu dengan yang lainnya namun sangat bertolak belakang dengan data pada cluster lain. Tergantung pada data dan aplikasi, berbagai jenis ukuran kesamaan dapat dipakai untuk mengidentifikasi kelas, dimana ukuran kesamaan mengontrol suatu cara cluster itu terbentuk. Beberapa contoh nilai yang dapat digunakan sebagai parameter keserupaan termasuk jarak, konektivitas serta intensitas [3].

### B.6. Algoritma K-Means

K-Means adalah metode dalam clustering atau pengelompokan. Clustering membentuk sebuah pengelompokan data, observasi berdasarkan kemiripan sebuah objek penelitian. Kluster adalah sebuah kelompok pada kumpulan data yang sangat mirip terhadap data lainnya [5]. Clustering dijabarkan oleh (Xu Rui, 2009) yang berarti membagi sebuah objek data atau item (entitas, contoh, unit) terhadap kelompok objek tertentu. Adapun tujuan dari tahap-tahap pemrosesan clustering yang dijabarkan oleh (Agusta, 2007) yakni meminimalkan objective function yang ditetapkan dalam proses clustering, biasanya digunakan untuk meminimalkan variasi dalam sebuah cluster

### B.7. Algoritma FP-Growth

Algoritma FP-Growth adalah algoritma aturan asosiasi yang sering dipakai. Dalam mencari kumpulan item yang seringkali bermunculan (frequent itemset) pada sebuah kumpulan data. FP-Growth menguntungkan dikarenakan selesai satu ataupun duakali saja dalam pemindaian basis datanya. Pada aturan apriori akurasi rulanya lebih akurat daripada FP-Growth. Akan tetapi, dikarenakan scanning dilakukan secara berulang kali, kecepatannya menjadikan lebih lambat dalam pengolahan datanya. Akan tetapi kedua metode algoritma yang digunakan ini mempunyai hal dengan fungsinya yang sama yaitu

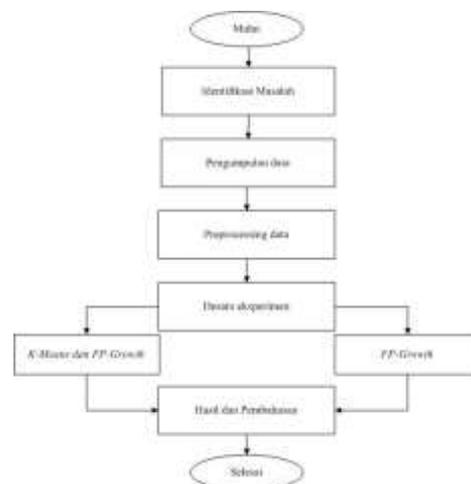
mendapatkan frequent itemset yang sering muncul [2].

### B.8. Rapid Miner

Menurut (Luluk E., 2017) Rapid Miner antara lain sebuah perangkat lunak yang bersifat bebas dan bisa digunakan oleh siapa saja (open source). Rapid Miner adalah sebuah penyelesaian dalam melakukan penganalisaan terhadap data mining, text mining dan analisis prediksi. Rapid Miner menggunakan teknik deskriptif dan prediktif dalam memberikan wawasan kepada pengguna maka dapat digunakan dalam membuat keputusan terbaik. Rapid Miner mempunyai sekitar 500 operasi data mining, termasuk operator untuk input, output, pra-proses data serta visualisasi. Rapid Miner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. Rapid Miner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik. Rapid Miner didistribusikan di bawah lisensi dengan nama AGPL (GNU Affero General Public License) tipe 3. Rapid Miner tersedia GUI (Graphic User Interface) untuk mendesain pipeline analisis. GUI akan mendapatkan hasil berupa file XML (Extensible Markup Language) dalam menentukan tahapan analisis yang ingin digunakan dan diterapkan.[7]

## C. METODOLOGI PENELITIAN

Adapun alur penelitian ini dijelaskan dalam metode penelitian seperti pada gambar 2.



Gambar 2. Metode Penelitian

### C.1. Tahap Identifikasi Masalah

Tahap identifikasi masalah adalah tahap pertama dalam mempelajari tentang permasalahan penelitian. Tahap identifikasi masalah terdiri dari:

1. Menentukan Tujuan Penelitian
2. Identifikasi Masalah
3. Menentukan Ruang Lingkup Dan Latar Belakang Dari Topik Penelitian
4. Studi Pustaka

### C.2. Tahap Pengumpulan Data

Dalam tahap pengumpulan data dilakukan dengan cara yaitu wawancara dan observasi. Dalam penelitian ini penulis melakukan pengumpulan data dengan melakukan wawancara dengan salah satu aparaturnya kepolisian yang terkait mengenai data kecelakaan lalu lintas, dan melakukan pengamatan secara langsung terhadap objek penelitian di unit Laka Polres Solok. Dari hasil wawancara diperoleh permasalahan kecelakaan lalu lintas yang terjadi di Kab. Solok dan data kejadian kecelakaan lalu lintas tahun 2015-2018 sebanyak 497.

### C.3. Tahap Preprocessing Data

Dalam tahap *praprocessing* ini terdapat tiga tahapan yang akan dilalui sebagai berikut:

1. *Data Selection*
2. *Data Cleaning*
3. *Data Transformation*

### C.4. Tahap Desain Eksperimen

Adapun kegiatan yang dilakukan pada tahap ini adalah:

1. Proses Algoritma K-Means dan FP-Growth  
 Proses Algoritma K-Means digunakan untuk proses clustering dan FP-Growth sebagai pendekatan asosiasi terhadap tiap cluster. dilakukan menggunakan aplikasi alat bantu Rapid Miner 8.0, Pada dataset ini di cluster menggunakan algoritma K-Means dengan ketentuan cluster ( $k$ ) = 2. Dalam tahap asosiasi akan diukur nilai support, confidence dan lift rasionya. Kemudian dilakukan eksperimen menggunakan algoritma FP-Growth tanpa melalui proses clustering. Selanjutnya membandingkan 2 eksperimen dan menyimpulkan hasil nilai akurasi support, confidence dan lift ratio yang dimana lebih efektif digunakan terhadap pengolahan data kecelakaan lalu lintas pada Unit Laka Polres Solok.

### C.5. Tahap Hasil Dan Pembahasan

Tahap hasil dan pembahasan dalam penelitian ini berbentuk pola informasi yang dihasilkan dari perbandingan eksperimen 2 proses yaitu pertama, dengan dikluster terlebih dahulu menggunakan algoritma K-Means selanjutnya dari masing-masing hasil cluster di asosiasikan menggunakan algoritma FP-Growth dan kedua tidak menggunakan proses clustering melainkan langsung pada proses pencarian asosiasinya. Dari rule yang didapat hasil rule yang terbaik dengan nilai akurasi support yang efektif akan diambil rule nya. Tahap dokumentasi dilakukan dari awal penelitian hingga akhir penelitian.

## D. HASIL DAN PEMBAHASAN

### D.1. Pengumpulan data *Knowledge Discovery in Database (KDD)*

Data kecelakaan lalu lintas yang terkumpul adalah 497 kejadian dan tercatat sebanyak 500 data

pelaku lakalantas. Setelah praproses data menjadi 461 data kecelakaan lalu lintas. Data kecelakaan lalu lintas dapat dilihat pada tabel 1.

Tabel 1. Data kecelakaan lalu lintas 2015-2018.

No	Waktu	SIM	...	Kondisi
1	Jan-3-2015 19.00	Jam SIM A	...	LR
2	Jan-4-2015 10.00	Jam SIM A	...	LR
3	Jan-7-2015 20.00	Jam SIM A	...	LR
4	Jan-10-2015 10.30	Jam Tdk ada SIM	...	LR
5	Jan-19-2015 15.00	Jam Tdk ada SIM	...	LR
....	.....	....	....	....
461	Jan-27-2018 11.30	Jam BII Umum	....	LR

Tabel 2. Boolean Data Kecelakaan Lalu Lintas 2015-2018

LAKA	A1	A2	B1	B2	C1	C2	...	K14
1	1	0	1	0	1	0	...	0
2	1	0	1	0	0	0	...	0
3	0	1	0	1	0	0	...	0
4	0	1	0	1	0	0	...	0
5	1	0	0	1	1	0	...	0
....	..	..	..	...	..	...	...	..
461	1	0	0	1	0	1	..	0

### D.2. Proses Algoritma K-Means dan FP-Growth

Pada tahap ini data yang telah terkumpul dioleh melalui praproses data yang digunakan untuk proses algoritma K-Means dan FP-Growth pada tabel 1. Tahap selanjutnya melakukan percobaan perhitungan secara manual. Dalam percobaan perhitungan secara manual ini menggunakan 20 data sampel dikarenakan memakan waktu yang lama dan panjang, dimana dalam pengelolaan data secara keseluruhan menggunakan tools Rapid Miner 8.0. Adapun tujuan perhitungan ini adalah untuk menjelaskan secara detail bagaimana proses algoritma yang digunakan. Data yang digunakan dapat dilihat pada tabel 3.

Tabel 3. Data Sampel Kecelakaan Lalu Lintas

No	Waktu	SIM	...	Kondisi
1	Padat	Ada	...	LR
2	Padat	Ada	...	Luka Ringan
3	Sepi	Ada	...	Luka Ringan
4	Sepi	Tdk	...	Luka Ringan
5	Sepi	Tdk	...	Luka Ringan
...	...	...	...	...
20	Sepi	Ada	...	Luka Ringan

**1. Proses Clustering K-Means**

Sebelum melakukan pengolahan data, tentukan dahulu clusternya, dimana pada penelitian ini menggunakan 2 cluster(k). Hasil dari proses clustering dihasilkan dalam bentuk file terpisah tiap cluster. Setelah hasil cluster didapat yaitu cluster\_0 dan cluster\_1 kemudian diurutkan dan dipisah menjadi 2 file untuk dilakukan analisa data dengan proses asosiasi menggunakan FP-Growth.

**2. Proses Cluster 0 dengan FP-Growth**

Tahap awal adalah menentukan nilai support. Salah satu contoh mencari nilai support item adalah sebagai berikut :

$$\text{Support}(B2,G1) = 9/14 \times 100 \% = 64\%$$

Untuk lebih jelas hasil perhitungan support pada data kecelakaan dapat dilihat pada tabel 4.

Tabel 4. Daftar nilai support item dengan nilai minimum support 60 %.

Item	Jumlah	Support
F1-G1	12	86%
B2-G1	9	64%
G1,D2-F1	10	71 %
F1,D2-G1	10	71 %
G1,A1-F1	9	64 %
F1,A1-G1	9	64 %

Setelah nilai semua nilai support ditemukan, baru dicari aturan asosiasi, jika nilai hasil yang dibentuk lebih besar nilai confidence yang dihasilkan pada subset yang terbentuk lebih besar dari nilai minimum confidence maka Association Rule tersebut dapat digunakan. Nilai confidence yang digunakan 1,0 , maka association rule yang memenuhi syarat adalah

$$\text{Confidence}(F1,G1) = \frac{14}{14} \times 100 \% = 100\%$$

Untuk lebih jelas hasil perhitungan support pada data kecelakaan dapat dilihat pada tabel 5.

Tabel 5. Daftar nilai confidence item dengan nilai minimum confidence 100 %.

Item	Jumlah	Confidence
F1-G1	14	100%
B2-G1	14	100 %

G1,D2-F1	14	100 %
F1,D2-G1	14	100 %
G1,A1-F1	14	100%
F1,A1-G1	14	100%

Setelah didapatkan hasil nilai support dan confidence, maka selanjutnya dilakukan analisa menggunakan bantuan tools Rapid Miner Studio 8.0 menggunakan cluster\_0 dengan nilai minimum support 0.6(60%) dan confidence 1.0 (100%) maka diperoleh rule yang dapat dilihat pada tabel 6.

Tabel 6. Daftar hasil rule pada rapidminer

No	Jika	Maka	Supp	Conf	Lift Ratio
1	F1	G1	0.86	1	1.077
2	B2	G1	0.64	1	1.077
3	G1,D2	F1	0.71	1	1.157
4	F1,D2	G1	0.71	1	1.077
5	G1,A1	F1	0.64	1	1.157
6	F1,A1	G1	0.64	1	1.077

Berikut penjelasan dari rule pada tabel 6.

1. If pengendara berjenis kelamin laki-laki then mengalami luka ringan dengan tingkat kepercayaan 100% dan didukung 86% dari data keseluruhan dengan nilai lift ratio 1.077 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
2. If kondisi jalan dalam keadaan sepi then mengalami luka ringan dengan tingkat kepercayaan 100% dan didukung 64% dari data keseluruhan dengan nilai lift ratio 1.077 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
3. If pengendara mengalami luka ringan dan berusia muda then pelaku berjenis kelamin laki-laki dengan tingkat kepercayaan 100% dan didukung 71% dari data keseluruhan dengan nilai lift ratio 1.157 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
4. If pengendara berjenis kelamin laki-laki dan berusia muda then mengalami luka ringan dengan tingkat kepercayaan 100% dan didukung 71% dari data keseluruhan dengan nilai lift ratio 1.077 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
5. If pengendara mengalami luka ringan dan memiliki SIM then pengendara berjenis kelamin laki-laki dengan tingkat kepercayaan 100% dan didukung 64% dari data keseluruhan dengan nilai lift ratio 1.167 menunjukkan bahwasanya rule ini

memiliki manfaat dan kekuatan asosiasi yang tinggi.

- If pengendara berjenis kelamin laki-laki dan memiliki SIM then pengendara mengalami luka ringan dengan tingkat kepercayaan 100% dan didukung 64% dari data keseluruhan dengan nilai lift ratio 1.077 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.

### 3. Proses Cluster 1 dengan FP-Growth

Tahap awal adalah menentukan nilai support. Salah satu contoh mencari nilai support item adalah sebagai berikut :

$$\text{Support}(K6,B1) = \frac{5}{6} \times 100\% = 83\%$$

Untuk lebih jelas hasil perhitungan support pada data kecelakaan dapat dilihat pada tabel 7.

Tabel 7. Daftar nilai support item dengan nilai minimum support 60 %.

Item	Jumlah	Support
B1-A2	6	100%
A2-B1	6	100 %
K6-B1	5	83 %
G1-B1	5	83 %
K6-A2	5	83 %
G1-A2	5	83 %
K6-B1,A2	5	83 %
B1,K6-A2	5	83 %
A2,K6-B1	5	83 %
G1-B1,A2	5	83 %
B1,G1-A2	5	83 %
A2,G1-B1	5	83 %

Setelah nilai semua nilai support ditemukan, baru dicari aturan asosiasi, jika nilai hasil yang dibentuk lebih besar nilai confidence yang dihasilkan pada subset yang terbentuk lebih besar dari nilai minimum confidence maka Association Rule tersebut dapat digunakan. Nilai confidence yang digunakan 1,0, maka association rule yang memenuhi syarat adalah

$$\text{Confidence}(B1,A2) = \frac{14}{14} \times 100\% = 100\%$$

Untuk lebih jelas hasil perhitungan support pada data kecelakaan dapat dilihat pada tabel 8.

Tabel 8. Daftar nilai confidence item dengan nilai minimum confidence 100 %.

Item	Jumlah	Confidence
B1-A2	14/14	100%
A2-B1	14/14	100 %
K6-B1	14/14	100 %
G1-B1	14/14	100 %
K6-A2	14/14	100%
G1-A2	14/14	100%

K6-B1,A2	14/14	100%
B1,K6-A2	14/14	100%
A2,K6-B1	14/14	100%
G1-B1,A2	14/14	100%
B1,G1-A2	14/14	100%
A2,G1-B1	14/14	100%

Setelah didapatkan hasil nilai support dan confidence, maka selanjutnya dilakukan analisa menggunakan bantuan tools Rapid Miner Studio 8.0 menggunakan cluster\_1 dengan nilai minimum support 0.6(60%) dan confidence 1.0 (100%) maka diperoleh rule yang dapat dilihat pada tabel 9.

Tabel 9. Daftar hasil rule pada rapidminer

No	Jika	Maka	Supp	Conf	Lift Ratio
1	B1	A2	1	1	1
2	A2	B1	1	1	1
3	K6	B1	0.83	1	1
4	G1	B1	0.83	1	1
5	K6	A2	0.83	1	1
6	G1	A2	0.83	1	1
7	K6	B1-A2	0.83	1	1
8	B1,K6	A2	0.83	1	1
9	A2,K6	B1	0.83	1	1
10	G1	B1,A2	0.83	1	1
11	B1,G1	A2	0.83	1	1
12	A2,G1	B1	0.83	1	1

Berikut penjelasan dari rule pada tabel 9.

- If Kondisi jalan dalam keadaan padat then pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 100% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
- If Pengendara tidak memiliki SIM then Kondisi jalan dalam keadaan padat dengan tingkat kepercayaan 100% dan didukung 100% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
- If Kecelakaan terjadi di kecamatan Gunung Talang then kondisi jalan dalam keadaan padat dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
- If Pengendara mengalami luka ringan then kondisi jalan dalam keadaan padat dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini

- memiliki manfaat dan kekuatan asosiasi yang tinggi.
5. If Kecelakaan terjadi di kecamatan Gunung Talang then pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
  6. If Pengendara mengalami luka ringan then pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
  7. If Kecelakaan terjadi di kecamatan Gunung Talang then kondisi jalan dalam keadaan padat dan pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
  8. If Kondisi jalan dalam keadaan padat dan kecelakaan terjadi di kecamatan Gunung Talang then pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
  9. If pengendara tidak memiliki SIM dan Kecelakaan terjadi di kecamatan Gunung Talang then kondisi jalan dalam keadaan padat dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
  10. If pengendara mengalami luka ringan then kondisi jalan dalam keadaan padat dan pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.
  11. If kondisi jalan dalam keadaan padat dan mengalami luka ringan then pengendara tidak memiliki SIM dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.

12. If pengendara tidak memiliki SIM dan mengalami luka ringan then kondisi jalan dalam keadaan padat dengan tingkat kepercayaan 100% dan didukung 83% dari data keseluruhan dengan nilai lift ratio 1 menunjukkan bahwasanya rule ini memiliki manfaat dan kekuatan asosiasi yang tinggi.

### D.3. Penerapan Algoritma K-Means dan FP-Growth Pada Semua Data Tahun 2015-2018

Pada Pada tahap ini, data kecelakaan lalu lintas selama 4 tahun diolah menggunakan bantuan tools Rapid Miner 8.0. Untuk menentukan nilai support dan confidence peneliti melakukan percobaan terhadap data keseluruhan dengan nilai support dan confidence berbeda mulai dari nilai support terendah dan nilai confidence tertinggi, sehingga ditemukan nilai support dan confidence yang efektif.

#### D.3.1 Tahap Pencarian Nilai Support dan Confidence

Pada tahap pengujian, analisis dilakukan dengan menggunakan cluster\_0, penulis melakukan pengujian semua data yang didapatkan selama 4 tahun dengan nilai minimum support 40%, 50%, dan 60%, sedangkan untuk minimum confidence dengan nilai 60%, 70%, 80% dan 100%. Pengujian ini menggunakan data sebanyak 237 data dan atribut sebanyak 47 atribut. Pengujian dilakukan sebanyak 12 kali dengan nilai support dan confidence yang berbeda disetiap pengujiannya sesuai dengan nilai support dan confidence yang telah ditentukan.



Gambar 3. Grafik Hasil Pengujian Perbandingan Rule.

Berdasarkan pengujian diatas, dapat dilihat bahwa nilai support 0.4 dan confidence 60% tidak efektif untuk hasil analisis, dalam hal ini peneliti menentukan nilai support yang efektif adalah minimum support 0.6 dan minimum confidence 100%. Dilihat dari hasil rule yang dihasilkan dengan nilai confidence 100% dan minimum nilai support 0.6 dapat disimpulkan bahwa nilai rule yang ditemukan dapat menjelaskan hasil yang kuat dengan menghasilkan rule yang paling dominan dalam beberapa variable dari data kecelakaan lalu

lintas dan yang dihasilkan tidak terlalu banyak dan tidak pula sedikit dalam mencari hasil rulenya.

Pada tahap selanjutnya melakukan analisis pada cluster\_1, penulis melakukan pengujian semua data yang didapatkan selama 4 tahun dengan nilai minimum support 40%, 50%, dan 60%, sedangkan untuk minimum confidence dengan nilai 60%, 70%, 80% dan 100%. Pengujian ini menggunakan data sebanyak 224 data dan atribut sebanyak 47 atribut. Pengujian dilakukan sebanyak 12 kali dengan nilai support dan confidence yang berbeda.



Gambar 4. Grafik Hasil Pengujian Perbandingan Rule.

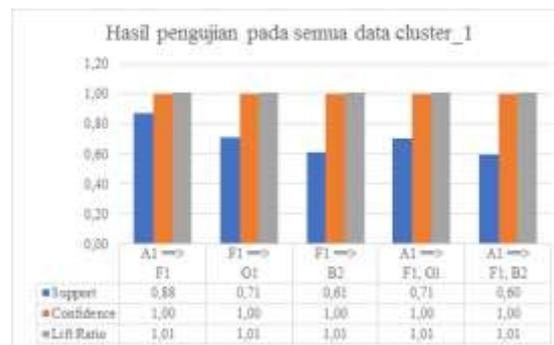
### D.3.2 Tahap Analisa Hasil Semua Data Pada Tahun 2015-2018

Tahap analisa hasil rule yang didapatkan dari nilai *support* dan *confidence* yang didapatkan sebelumnya terhadap semua data sebanyak 461, dengan nilai *support* 5% dan *confidence* 30%. Pada cluster 0 terdapat 250 data dengan nilai *support* 30% dan *confidence* 90%. Berikut hasil analisa dari hasil rule yang didapatkan dapat dilihat pada Gambar 5.



Gambar 5. Grafik Hasil Pengujian Pada Semua Data Cluster 0 Tahun 2015-2018

Pada cluster 1 terdapat 224 data dengan nilai *support* 60% dan *confidence* 100%. Berikut hasil analisa dari hasil rule yang didapatkan dapat dilihat pada Gambar 6.



Gambar 6. Grafik Hasil Pengujian Pada Semua Data Cluster 1 Tahun 2015-2018

### D.3.3 Tahap Analisa Hasil Semua Data Menggunakan Algoritma FP-Growth Tahun 2015-2018

Pada tahap ini, data kecelakaan lalu lintas selama 4 tahun diolah menggunakan bantuan tools Rapidminer 8.0. Untuk menentukan nilai *support* dan *confidence* peneliti melakukan percobaan terhadap data keseluruhan dengan nilai *support* dan *confidence* berbeda mulai dari nilai *support* tertinggi dan nilai *confidence* terendah, sehingga ditemukan nilai *support* dan *confidence* yang efektif terhadap data kecelakaan lalu lintas di Polres Solok. Pada tahapan ini dilakukan proses pengolahan data yang berjumlah 461 data kecelakaan.

Pengujian dilakukan sebanyak 9 kali dengan nilai *support* minimum 30%, 40%, dan 50%, sedangkan untuk minimum *confidence* dengan nilai 60%, 80% dan 90%. Pengujian ini menggunakan data 461 data dan atribut sebanyak 47.



Gambar 7. Grafik Hasil Pengujian Pada Semua Data Tahun 2015-2018

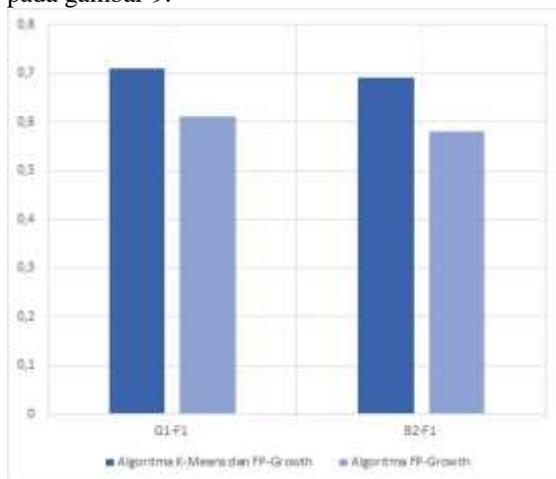
Tahap selanjutnya adalah menganalisa hasil rule yang didapat dari nilai *support* dan *confidence* yang didapat sebelumnya terhadap semua data sebanyak 461 dengan nilai *support* 0,5 dan *confidence* 80%. Berikut analisa dari hasil rule yang didapatkan dapat dilihat pada Gambar 8.



Gambar 8. Grafik Hasil Pengujian Pada Semua Data Tahun 2015-2018

#### D.4 Hasil Perbandingan Rule dari 2 Eksperimen

Tahap ini membandingkan hasil aturan yang sama yang didapatkan dengan menggunakan algoritma K-Means dan FP-Growth dan Algoritma FP-Growth saja. Tujuannya adalah untuk mengetahui nilai akurasi support manakah yang paling efektif dan lebih besar yang didapat agar hasil rule menjadi lebih efektif. Grafik hasil dapat dilihat pada gambar 9.



Gambar 9. Grafik Hasil Perbandingan Rule Yang Sama Antara 2 Eksperimen Pada Semua Data Tahun 2015-2018

Dari hasil perbandingan rule pada Gambar 9, dapat dilihat bahwa pada rule if pengendara mengalami G1 (luka ringan) then pengendara dengan kategori F1 (laki-laki) dengan nilai akurasi support tertinggi 71% menggunakan Algoritma K-Means dan FP-Growth sedangkan menggunakan Algoritma FP-Growth saja hanya 69%.

#### E. KESIMPULAN

Berdasarkan hasil proses Algoritma K-Means dan FP-Growth pada data kecelakaan lalu lintas di Polres Solok dapat disimpulkan.

1. Nilai minimal support dan minimal confidence yang digunakan pada data kecelakaan lalu lintas pada Polres Solok dengan menggunakan algoritma K-Means dan FP-Growth adalah pada cluster 0 60% dan 100%, pada cluster 1 60% dan 100%, sedangkan pada algoritma FP-Growth 50% dan 80%.

2. Dari hasil penelitian pengolahan data pada semua data kecelakaan dari tahun 2015 sampai tahun 2018 dimana dapat disimpulkan bahwa kecelakaan lalu lintas yang paling dominan terjadi pada usia muda dengan rentang usia 16-35 tahun, pengendara pada jenis kelamin laki-laki rentan terhadap kecelakaan lalu lintas, kondisi waktu saat kecelakaan dalam keadaan sepi, kecelakaan sering terjadi pada jenis kendaraan sepeda motor, kecelakaan sering terjadi pada pengendara yang tidak memiliki SIM, jenis luka yang dialami pengendara dengan keadaan luka ringan.

3. Dengan adanya perbandingan dari 2 eksperimen yaitu eksperimen I (Algoritma K-Means dan FP-Growth) dan Eksperimen II (Algoritma FP-Growth saja) dapat diketahui nilai support tertinggi yaitu dengan menggunakan algoritma K-Means dan FP-Growth dengan rule yang didapat, if Kecelakaan dengan kategori B2(Waktu kecelakaan dalam keadaan sepi) then pengendara dengan kategori F1(Berjenis kelamin laki-laki) dengan nilai support menggunakan algoritma K-Means dan FP-Growth lebih tinggi yaitu 61% dibandingkan dengan menggunakan FP-Growth saja didapat nilai support 58%.

4. Dari hasil penelitian ini, dapat memberikan rekomendasi pada Pihak Polres Solok khususnya unit laka, bahwasanya dalam mengurangi kejadian kecelakaan lalu lintas dapat dilakukan dengan cara melaksanakan sosialisasi dan pengarahan ke sekolah maupun umum yang berkaitan dengan laka, dan juga dapat dilakukan dengan pembuatan traffic atau pembuatan marka jalan yang sesuai dengan arah kendaraan dengan melaporkan pada dinas terkait.

#### REFERENSI

- [1] Agusta. (2007). *K-means, penerapan, permasalahan dan metode terkait*. Jurnal Sistem dan Informatika.
- [2] Fitria, R., Nengsih, W., & Qudsi, D. H. (2017). Implementasi Algoritma FP-Growth Dalam Penentuan Pola Hubungan Kecelakaan Lalu Lintas. *Jurnal Sistem Informasi*, 13(2), 118. <https://doi.org/10.21609/jsi.v13i2.551>
- [3] Hardiyanto, I., Purwananto, Y., Kom, S., Kom, M., & Soelaiman, R. (2012). Implementasi Segmentasi Citra dengan Menggunakan Metode Generalized Fuzzy C-Means Clustering Algorithm with Improved Fuzzy Partitions. *Teknik Pomits*, 1(1), 1–5.
- [4] Ikhwan, A., Nofriansyah, D., & Sriani. (2015). Penerapan Data Mining dengan Algoritma Fp-Growth untuk Mendukung Strategi Promosi Pendidikan ( Studi Kasus

- Kampus STMIK Triguna Dharma ).  
*Saintikom*, 14(3), 211–226.
- [5] Larose. (2005). *Discovering knowledge in data : An introduction to data mining* (New Jersey:John Wiley (ed.)). Sons Inc.
- [6] Lestari. (2015). Implementasi pengolahan citra dan klasifikasi k-nearest neighbour untuk membangun aplikasi pembeda daging sapi dan babi berbasis web. *Jurnal Sains Dan Teknologi Industri*, 242–247.
- [7] Luluk Elvitaria, M. H. (2017). Smk Analisis Kesehatan Abdurrah Menggunakan Algoritma. (*Jurnal Teknologi Dan Sistem Informasi Univrab*, 2(2), 220–233.
- [8] Mamahit, N., & Qoiriah, A. (2019). Penerapan Algoritma Fp-Growth dan K-Means pada Data Transaksi Minimarket. *Journal of Informatics and Computer Science (JINACS)*, 1(02), 78–83.
- [9] Maskuroh. (2014). Analisa keranjang belanja dengan aturan asosiasi menggunakan algoritma apriori untuk stok barang pada listrikmart. *Jurnal Informatika*.
- [10] Mujiasih, S. (2011). Pemanfaatan Data Mining Untuk Prakiraan Cuaca. *Jurnal Meteorologi Dan Geofisika*, 12(2), 189–195.  
<https://doi.org/10.31172/jmg.v12i2.100>
- [11] Mulyani, E. D. S. dkk. (2018). Implementasi Algoritma K-Means Dan Fp- Growth Untuk Rekomendasi Bimbingan Belajar Berdasarkan Segmentasi Akademik Siswa. *IT Journal*, 6(2), 160–173.
- [12] Risianingrum. (2017). Implementasi data mining menggunakan algoritma apriori. *Prosiding SINTAK*, 372–382.
- [13] Rui, X. (2009). *Clustering* (J. Wiley (ed.)). Sons Inc.
- [14] Saragih, R. (2017). *Implementasi Apriori Pada Data Kecelakaan Lalu*. 6341(November), 12–17.
- [15] Sepri, D., & Afdal, M. (2017). Analisa Dan Perbandingan Metode Algoritma Apriori Dan Fp-Growth Untuk Mencari Pola Daerah Strategis. *Jurnal Sistem Informasi Kaputama (JSIK)*, 1(1), 47–55.
- [16] Winda Aprianti, J. P. (2018). K-means clustering untuk data kecelakaan lalu lintas jalan raya di kecamatan pelaihari. *Jurnal Teknologi Informasi dan Ilmu Komputer(JTIK)*, 5(5), 613–620.