

INTISARI

PENGGUNAAN ASSOCIATION RULE MINING ALGORITME CT-PRO DALAM PROSES PENENTUAN POLA TINDAK KEJAHATAN PERDAGANGAN MANUSIA (Studi Kasus: Data Kepolisian Daerah Jawa Barat)

Oleh
Nur Fitriyah Ayu Tunjung Sari
14/372004/PPA/04124

Berdasarkan data tahun 2015, Indonesia merupakan negara dengan Tindak Pidana Perdagangan Orang (TPPO) nomor tiga terbanyak di dunia. Hal tersebut menjadikan kejahatan ini membutuhkan perhatian khusus dari banyak pihak, salah satunya kepolisian. Salah satu cara yang dapat dilakukan kepolisian untuk menekan terjadinya kejahatan tersebut adalah dengan menganalisa data kejahatan. Salah satu teknik yang dapat digunakan untuk menganalisa adalah *CT-PRO association rule mining*.

Hasil luaran penelitian ini adalah aturan yang divisualisasikan dalam bentuk visualisasi berbasis matriks yang dapat digunakan sebagai pengetahuan atau informasi baru untuk menentukan pola kejahatan oleh kepolisian. Pada penelitian dilakukan pengujian validitas melalui pengambilan angket ke Polda Jawa Barat. Pengujian juga dilakukan dari sisi sistem sebanyak 4 kali pengujian dengan berbagai nilai *minimum support* yaitu, 0,02; 0,05; 0,1 dan 0,2 dan *minimum confidence* sebesar 0,01 untuk melihat pola kejahatan yang dapat ditentukan dari aturan dan visualisasi yang dihasilkan sistem. Selanjutnya, pengujian dilakukan dengan membandingkan aturan yang dihasilkan oleh algoritme *CT-PRO* dengan algoritme apriori.

Hasil penelitian menunjukkan bahwa aturan dan visualisasi yang dihasilkan dapat mewakili kejadian yang ada untuk membantu kepolisian menentukan pola terjadinya kejahatan perdagangan manusia. Pada tahun 2010-2015, pola tindak kejahatan perdagangan manusia paling banyak terjadi di Indramayu dengan korban berjenis kelamin perempuan. Sedangkan pencarian aturan dengan algoritme apriori menghasilkan aturan yang sama dengan algoritme *CT-PRO* pada *minimum support* 0,1 dan 0,2. Dari segi kecepatan, algoritme *CT-PRO* lebih cepat rata-rata 5,352993 kali jika dibandingkan dengan algoritme apriori.

Kata kunci: *association rule mining*, *CT-PRO*, perdagangan manusia, visualisasi matriks

ABSTRACT

IMPLEMENTATION OF CT-PRO ASSOCIATION RULE MINING ALGORITHM FOR DETERMINATION OF HUMAN TRAFFICKING PATTERN (Case Study: Police Data in West Java District)

By

Nur Fitriyah Ayu Tunjung Sari
14/372004/PPA/04124

Based on data in 2015, Indonesia is the country with the crime of trafficking in person number three in the world. It makes this crime need special attention from many sides, one of them is the police. One of the ways the police can press the occurrence of human trafficking is analyze the data crime. One of the technique that can be used to the analyze is CT-PRO association rule mining.

The output of this research is a rule and visualization of rule based on matrix visualization that can be used as knowledge or new information to determine the patterns of crime by the police. Implementation of CT-PRO algorithms association rule mining has verified for validity by making a questionnaire to the West Java Regional Police to see if the rules are generated to represent the actual events on the reality. Testing also performed in the system with the various value of minimum support; that is 0.02; 0.05; 0.1 and 0.2 and minimum confidence 0.01 to see the pattern of the crimes that can be determined from the resulting rules and visualization. Furthermore, testing also by comparing the rules generated by the CT-PRO algorithm with the apriori algorithm.

The results showed that the resulting rules and visualization could represent events that are there to help police determine the pattern of the crime of human trafficking. In 2010-2015, the pattern of the crime of human trafficking was most intense in Indramayu with the victims is female. While generating rule using the apriori algorithm give the same result as generate rule using a CT-PRO algorithm on the minimum support of 0.1 and 0.2. The process of generating rules by using CT-PRO algorithm is faster on average 5.352993 times compared with the apriori algorithm.

Keywords: association rule mining, CT-PRO, human trafficking, matriks of visualization

BAB I PENDAHULUAN

1.1 Latar Belakang

Laju pertumbuhan dan perkembangan penduduk yang terjadi saat ini diiringi dengan meningkatnya kebutuhan hidup. Salah satu kebutuhan hidup tersebut adalah kebutuhan akan lapangan pekerjaan. Ketersediaan lapangan pekerjaan yang tidak sesuai dengan laju pertumbuhan penduduk mengakibatkan timbulnya berbagai macam tindak kejahatan. Disisi lain kondisi perekonomian yang tidak merata mengakibatkan kesenjangan kehidupan yang sampai menyentuh akan kesenjangan pendidikan. Kondisi perekonomian yang kurang baik, lapangan pekerjaan yang tidak sesuai dengan laju pertumbuhan penduduk, kesenjangan pendidikan mengakibatkan dorongan yang sangat kuat akan munculnya kriminalitas.

Minimnya lapangan pekerjaan, rendahnya pendidikan yang ditempuh serta minimnya pengalaman mendesak seseorang untuk mudah mempercayai tawaran-tawaran pekerjaan yang sifatnya bisa jadi sebuah modus kejahatan. Tindak kejahatan yang sering muncul dengan modus tawaran pekerjaan salah satunya adalah tindak kejahatan perdagangan manusia. Karena desakan ekonomi yang muncul mengakibatkan seseorang pencari kerja kurang mewaspadaai kejahatan tersebut dengan modus tawaran pekerjaan dengan iming-iming gaji yang tinggi dan pekerjaan yang baik.

Perempuan dan anak perempuan di bawah umur sangat rentan terhadap kasus eksploitasi dan perdagangan manusia. Sarana eksploitasi yang sering digunakan adalah berupa ancaman, penyalahgunaan otoritas, jeratan hutang, perkawinan, penahanan dan pemerkosaan. Kasus perdagangan perempuan dengan modus prostitusi di luar negeri adalah kasus yang paling umum terjadi. Berdasarkan data mahkamah agung, Indonesia merupakan negara dengan Tindak Pidana Perdagangan Orang (TPPO) nomor tiga terbanyak di dunia (Anwar, 2015).

Berdasarkan hasil wawancara dengan Kanit II DITSERSE Polda Jabar diperoleh informasi bahwa, di dalam Subdit IV selama ini sudah dilakukan pencatatan dan pencarian informasi pada tindak kejahatan perdagangan manusia

untuk memenuhi kebutuhan tertentu setiap tahunnya. Pencatatan data dan pencarian informasi pada tindak kejahatan perdagangan manusia oleh Subdit IV ditunjukkan dalam Tabel 1.1.

Tabel 1.1 Pencatatan data dan pencarian informasi tindak kejahatan perdagangan manusia oleh Subdit IV

No	Tahun	Jumlah	Proses Sidik	Selesai
1	2010	33	18	15
2	2011	65	35	30
3	2012	24	13	11
4	2013	26	1	25
5	2014	15	9	6
6	2015	7	4	3
Jumlah		170	80	90

Untuk menekan angka tindak kejahatan perdagangan manusia, selain dilakukan pencatatan dan pencarian informasi juga dibutuhkan analisis yang mendalam. Kegiatan analisis tersebut masuk dalam wewenang badan intelijen kepolisian. Akan tetapi, badan intelijen kepolisian belum pernah melakukan analisis mendalam pada tindak kejahatan perdagangan manusia karena lebih memprioritaskan tindak kejahatan yang berkaitan dengan kestabilan negara.

Oleh karena itu, salah satu cara yang dapat dilakukan pihak kepolisian untuk mempermudah analisis pada data tindak kejahatan perdagangan manusia adalah dengan memanfaatkan salah satu teknik yang ada di dalam *data mining*. Teknik di dalam *data mining* tersebut adalah *association rule mining*. *Association rule mining* digunakan untuk melihat hubungan dari *item-item* data di dalam sebuah kumpulan data yang diharapkan dapat menghasilkan aturan yang dapat digunakan sebagai informasi baru.

Di dalam implementasi *association rule mining* terdapat banyak algoritme yang dapat digunakan seperti algoritme *apriori*, *FP-Growth* dan *CT-PRO*. Berdasarkan penelitian-penelitian sebelumnya *association rule mining* sering digunakan untuk melakukan analisis transaksi penjualan atau yang disebut dengan *market basket analysis* (Buulolo, 2013; Kurnia, dkk, 2013; Vanitha, 2011). Pada awalnya algoritme yang sering digunakan untuk menganalisis sekumpulan data dengan *association rule mining* adalah algoritme *apriori*. Seiring berkembangnya waktu muncul algoritme lain sebagai upaya untuk menutupi kekurangan yang ada

di dalam algoritme *apriori*. Algoritme tersebut adalah algoritme *FP-Growth* dan *CT-PRO* (Buulolo, 2013 dan Kurnia, dkk, 2013).

Di dalam penelitian ini mengimplementasikan algoritme *CT-PRO* pada *association rule mining*. Algoritme *CT-PRO* merupakan algoritme pengembangan dari *FP-Growth*. Alasan algoritme *CT-PRO* digunakan dalam penelitian ini karena mampu mengolah data baik dalam jumlah kecil maupun besar. Selain itu, algoritme *CT-PRO* memiliki kinerja yang lebih cepat dibandingkan dengan algoritme lainnya (Joshi, dkk, 2014).

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas maka rumusan masalah dalam penelitian ini adalah bagaimana menghasilkan aturan untuk memperoleh pengetahuan baru yang dapat digunakan untuk menetapkan pola terjadinya tindak kejahatan perdagangan manusia dan menampilkan aturan tersebut dalam bentuk visualisasi berbasis matriks.

1.3 Batasan Masalah

Permasalahan yang akan dibahas pada penelitian ini memiliki ruang lingkup yang cukup luas, sehingga diberikan batasan-batasan yang meliputi:

1. Data sampel yang digunakan adalah data primer yang diperoleh dari Polisi Daerah (POLDA) Jawa Barat.
2. Data yang digunakan meliputi data daerah asal korban, daerah tujuan perdagangan, jenis kelamin dan usia korban serta modus.
3. Analisis yang akan dilakukan pada data perdagangan manusia menggunakan salah satu teknik yang ada pada *data mining* yaitu *association rule mining* algoritme *CT-PRO* untuk menghasilkan aturan.
4. Pembahasan pada penelitian ini dibatasi pada bagaimana mengimplementasikan algoritme *CT-PRO association rule mining* yang diusulkan dapat menghasilkan aturan untuk menetapkan pola terjadinya tindak kejahatan perdagangan manusia.

1.4 Tujuan dan Manfaat Penelitian

Penelitian yang dilakukan bertujuan untuk menghasilkan aturan dari data kejahatan perdagangan manusia dengan mengimplementasikan algoritme *CT-PRO association rule mining*. Aturan yang dihasilkan akan ditampilkan dalam bentuk

visualisasi berbasis matriks untuk mempermudah penerimanya menemukan aturan yang menarik.

Aturan yang dihasilkan dapat dimanfaatkan sebagai pengetahuan baru oleh pihak kepolisian untuk menetapkan pola terjadinya kejahatan perdagangan manusia. Dengan diketahuinya pola terjadinya kejahatan perdagangan manusia diharapkan dapat membantu pihak kepolisian dalam merancang tindakan pencegahan atau sebagai bahan pertimbangan dalam pengambilan keputusan sehingga dapat menekan terjadinya kejahatan perdagangan manusia. Selain itu untuk membangun atau meningkatkan kewaspadaan dan sensitifitas kejahatan perdagangan manusia didalam lingkungan masyarakat dan pihak kepolisian.

1.5 Keaslian Penelitian

Penelitian dengan menggunakan *association rule mining* algoritme *CT-PRO* sudah pernah dilakukan namun berdasarkan referensi dan kajian pustaka yang dimiliki, penelitian yang diajukan sebagai Thesis S2 Ilmu Komputer Universitas Gadjah Mada Yogyakarta mengenai penggunaan *association rule mining* algoritme *CT-PRO* dalam proses penentuan pola tindak kejahatan perdagangan manusia belum pernah dilakukan.

1.6 Tahapan Penelitian

1) Studi Pustaka

Studi pustaka bertujuan untuk mencari dan meperoleh literatur-literatur yang berkaitan dengan perdagangan manusia, *association rule mining* dan juga algoritme yang akan digunakan dalam pencarian aturan yaitu *CT-PRO*. Literatur-literatur diperoleh dari buku, internet dan penelitian-penelitian terdahulu.

2) Pengumpulan Data

Data yang digunakan pada penelitian ini adalah data primer yang diperoleh dari pihak Polda Jabar. Pengumpulan data dilakukan dengan cara mengajukan permohonan data kepada instansi tersebut berkaitan dengan data perdagangan manusia.

3) Analisis dan Rancangan Sistem

Salah satu upaya yang dapat dilakukan pihak kepolisian untuk menekan terjadinya tindak kejahatan perdagangan manusia ialah dengan melakukan tindakan pencegahan. Upaya tersebut perlu didukung dengan sebuah informasi atau

pengetahuan yang diperoleh dari analisis pada data tindak kejahatan perdagangan manusia yang sebelumnya pernah terjadi. Analisis tersebut dapat dilakukan dengan mengumpulkan data-data dari tindak kejahatan dan dianalisis satu per satu oleh pihak kepolisian. Apabila proses tersebut dilakukan secara manual akan membutuhkan waktu yang lama sehingga untuk mempermudah dan meningkatkan waktu, proses analisis tersebut dapat dilakukan dengan mengimplementasikan sebuah teknik yang ada pada *data mining*, yaitu *association rule mining*. Untuk dapat mengimplementasikan *association rule mining* terdapat tahapan yang harus dilalui meliputi:

a. Analisis

Langkah awal yang dilakukan untuk memperoleh aturan dengan *association rule mining* yaitu menganalisis rancangan sistem untuk mengimplementasi algoritme *CT-PRO*. Sistem yang akan dibangun adalah sistem pencarian aturan tindak kejahatan perdagangan manusia dengan menggunakan algoritme *CT-PRO association rule mining* sehingga melalui aturan yang diperoleh dapat ditentukan pola terjadinya kejahatan perdagangan manusia. Sistem akan mengolah data yang disimpan di dalam *database*. Data diperoleh dari seleksi variabel pada data laporan tindak kejahatan perdagangan manusia yang dimiliki oleh pihak kepolisian.

Sistem akan melakukan seleksi *item* pada *database* sesuai dengan nilai *threshold* yang ditentukan sebelumnya. Besaran nilai *threshold* ditentukan berdasarkan masukan dari pihak kepolisian. Dari *item* yang memenuhi nilai *threshold* dibangun *global cfp-tree* yang selanjutnya dilakukan proses *mining* aturan yang dimulai dengan membangun lokal *cfp-tree* sehingga dihasilkan *frequent itemset*. Proses selanjutnya adalah menciptakan aturan dari setiap *frequent itemset*. Setelah aturan dihasilkan proses selanjutnya adalah menghitung nilai *support* dan *confidence* setiap aturan. Aturan yang memenuhi nilai *minimum support* dan *minimum confidence* membentuk *strong rule* pada data tindak kejahatan perdagangan manusia. Setiap *strong rule* yang dihasilkan selanjutnya akan ditampilkan dalam bentuk visualisasi berbasis matriks. Dari *strong rule* yang dihasilkan dan visualisasi yang diberikan dapat digunakan sebagai pengetahuan baru dalam menetapkan pola terjadinya tindak kejahatan perdagangan manusia.

b. Rancangan Sistem

Pada tahap perancangan sistem dilakukan dengan melakukan perancangan DFD. Perancangan DFD dilakukan untuk menggambarkan aliran data yang berjalan dalam sistem. Aliran data yang masuk hingga data yang keluar agar tergambar dengan jelas. Selain membuat rancangan DFD pada tahap ini juga dilakukan pembuatan rancangan desain antar muka. Desain antar muka dilakukan untuk dapat memberikan gambaran fungsionalitas sistem dari sisi pengguna yaitu pihak kepolisian.

4) Implementasi

Implementasi dari rancangan sistem dimulai dengan melakukan pengkodean menggunakan bahasa C# dan menggunakan *database mysql* sebagai tempat menyimpan data. *Tools* yang akan digunakan dalam pembuatan sistem adalah *unity 3D*. *Unity* adalah sebuah *tools* pengembangan yang memudahkan pengembang membuat program interaktif *multi platforms* serta program visualisasi 3D dan 2D.

5) Pengujian

Pengujian terhadap implementasi algoritme *CT-PRO association rule mining* yang diusulkan bertujuan untuk melihat apakah algoritme *CT-PRO* yang diimplementasikan mampu menghasilkan aturan yang sesuai dengan data kejahatan. Untuk melakukan pengujian validitas digunakan pengambilan angket kepada pihak Polda Jabar untuk melihat apakah aturan yang dihasilkan mampu mewakili kejadian yang sesungguhnya di lapangan. Pengujian juga dilakukan dengan membandingkan aturan yang dihasilkan oleh algoritme *CT-PRO* dengan algoritme apriori. Pengujian akan dilakukan sebanyak 4 kali proses dengan nilai *minimum support* 0,02; 0,05; 0,1 dan 0,2 sedangkan untuk nilai *minimum confidence* sebesar 0,01.

1.7 Sistematika Penulisan

Sistematika penulisan dilakukan untuk memberikan gambaran atau intisari setiap bab yang terdapat di dalam penelitian ini.

BAB I Pendahuluan

Bab ini menjelaskan secara singkat dari latar belakang penelitian yang disertai dengan rumusan masalah, batasan masalah, tujuan dan

manfaat penelitian, keaslian penelitian, tahapan penelitian dan sistematika penulisan.

BAB II Tinjauan Pustaka

Bab ini memberikan pemaparan secara singkat terkait hasil penelitian sejenis dari segi teknik maupun algoritme yang sudah dilakukan oleh peneliti terdahulu.

BAB III Landasan Teori

Bab ini memaparkan teori yang digunakan dalam penelitian sebagai dasar dalam pemecahan masalah.

BAB IV Analisis dan Rancangan Sistem

Bab ini memaparkan tahapan-tahapan yang dilakukan dalam penelitian yang dimulai dari analisis hingga melakukan perancangan sistem.

BAB V Implementasi

Bab ini menunjukkan program komputer yang dibangun berdasarkan rancangan yang sudah dibuat.

BAB VI Hasil dan Pembahasan

Bab ini berisikan hasil pengujian dan pembahasan terhadap hasil pengujian berdasarkan skenario pengujian.

BAB VII Kesimpulan dan Saran

Bab ini memaparkan kesimpulan dan saran yang diperoleh dari pengujian yang sudah dilakukan sebelumnya.

BAB II TINJAUAN PUSTAKA

Banyak penelitian telah dilakukan untuk menghasilkan aturan dengan menggunakan *association rule mining*. *Association rule mining* banyak dilakukan umumnya untuk melakukan analisis pada keranjang belanja di pasar swalayan. Misalnya (Buulolo, 2013) menggunakan *association rule mining* untuk menghasilkan pola kombinasi *itemsets* dan *rules* sebagai ilmu pengetahuan dan informasi penting dari data penjualan (data obat keluar). Penggunaan *association rule mining* selain digunakan untuk melakukan analisis terhadap data transaksi juga digunakan untuk melakukan analisis terhadap tindak kejahatan. Seperti pada penelitian (Atmajaya, 2016; Fadlina, 2014) yang memanfaatkan *association rule mining* untuk menghasilkan aturan pada data kejahatan.

Dalam melakukan pencarian aturan dengan *association rule mining* dapat dilakukan dengan menggunakan berbagai algoritme seperti *apriori* (Atmajaya, 2016; Buulolo, 2013; Fadlina, 2014), *FP-Growth* (Song, 2014) dan *CT-PRO* (Kurnia, dkk, 2013) untuk membangkitkan kombinasi *itemsets*. Salah satu implementasi algoritme apriori adalah pada penelitian Atmajaya (2016) yang menggunakan data tindak kejahatan. Implementasi algoritme apriori dilakukan untuk membangun sistem analisis pola tindak kejahatan.

Algoritme *CT-PRO* adalah pengembangan dari algoritme *FP-Growth* dimana algoritme *CT-PRO* mampu memperingkas dalam pembangunan *tree* pada algoritme *FP-Growth*. Seperti pada penelitian Kurnia, dkk (2013) dilakukan pencarian aturan dari *dataset* pembelian barang yang mampu menghasilkan pengetahuan terdiri dari *item* yang dibeli dengan item yang dibeli pula beserta nilai *support* dan *confidence*-nya. Sementara itu dari ketiga algoritme di atas (*apriori*, *FP-Growth* dan *CT-PRO*) dilakukan analisis performa untuk melihat perbandingan dalam penggunaannya. Seperti pada penelitian milik Vanitha, dkk (2011) melakukan evaluasi performa terhadap algoritme *apriori* dan *FP-Growth*. Dalam penelitiannya menunjukkan algoritme *apriori* membutuhkan waktu yang lebih lama dibandingkan dengan algoritme *FP-Growth* dalam mengolah sejumlah data.

Sedangkan dalam penelitian lain (Joshi, 2014) juga dilakukan perbandingan akan kemampuan kinerja beberapa variasi dari algoritme *fp-tree* seperti algoritme *FP-Growth*, *COFI-Tree* dan *CT-PRO*. Penelitian tersebut menyatakan bahwa *FP-Growth* adalah algoritme berbasis *tree* pertama yang berhasil melakukan *mining* terhadap data kriminal. Namun *FP-Growth* memiliki kekurangan disisi penggunaan memori di dalam membangkitkan *itemset*. Sehingga muncul variasi dari *FP-Growth* yaitu *COFI-Tree* dan *CT-PRO*.

Implementasi *association rule mining* pada sekumpulan data dapat menghasilkan aturan yang sangat banyak. Aturan yang dihasilkan pada umumnya akan ditampilkan dalam bentuk *if-then*. Hal tersebut mengakibatkan dibutuhkan waktu tambahan untuk dapat memahami sejumlah aturan tersebut. Visualisasi aturan diharapkan mampu memberikan kemudahan untuk memahami sejumlah aturan seperti pada penelitian (Atmajaya, 2016; Sekhavat, dkk, 2013). Pada penelitian Atmajaya (2016) aturan yang dihasilkan divisualisasi dalam bentuk *graph*. Pada penelitian lain yaitu penelitian oleh Sekhavat, dkk (2013) dikemukakan bahwa visualisasi dapat mempermudah pengguna untuk memahami data dengan cara eksplorasi, identifikasi dan intepretasi dari *association rules*. Pada penelitiannya menghubungkan tiga visualisasi *matrix view*, *graph view*, *detail view*. Dengan cara tersebut diharapkan pengguna dapat menemukan aturan yang menarik dengan lebih cepat dan mudah. Tabel 2.1 merangkum tinjauan pustaka yang diperoleh dalam penelitian ini.

Tabel 2.1 Tinjauan Pustaka

No	Peneliti	Teknik	Algoritme	Perbedaan Data Penelitian	Hasil Penelitian
1	Vanitha, dkk (2011)	<i>Association rule mining</i>	<i>Apriori</i> dan <i>FP-Growth</i>	<i>dataset</i> supermarket.	Menunjukkan bahwa <i>apriori</i> membutuhkan waktu yang lebih lama dibandingkan dengan <i>FP-Growth</i> dalam mengolah sejumlah data.
2	Song, dkk (2011)	<i>Association rule mining</i>	<i>FP-Growth</i>	data perpustakaan.	Menemukan hubungan yang menarik antara buku dengan pembaca sehingga dapat meningkatkan kualitas koleksi perpustakaan.
3	Buulolo (2013)	<i>Association rule mining</i>	<i>Apriori</i>	data transaksi penjualan (data obat keluar)	Menghasilkan pola kombinasi <i>items</i> dan <i>rules</i> sebagai ilmu pengetahuan dan informasi penting dari data penjualan (data obat keluar).
4	Kurnia, dkk (2013)	<i>Association rule mining</i>	<i>CT-PRO</i>	data pembelian barang.	Menghasilkan <i>knowledge</i> terdiri dari <i>item</i> yang dibeli dengan <i>item</i> yang dibeli pula beserta nilai <i>support</i> dan <i>confidence</i> -nya
5	Sekhavat, dkk (2013)	<i>Association rule mining</i>	<i>Linked Matrix</i> , <i>Graph</i> dan <i>Detail Views</i>	-	Menghasilkan tampilan visual dan interaktif pada aturan dimana pengguna akan dapat menemukan aturan penting dengan cepat dan mudah.
6	Fadlina (2014)	<i>Association rule mining</i>	<i>Apriori</i>	data daerah dan jenis kejahatan jalanan dari data statistik Polsek Medan Sunggal.	Menghasilkan informasi mengantisipasi kejahatan yang sering muncul.
7	Joshi, dkk (2014)	<i>Association rule mining</i>	<i>FP-Tree</i>	data tindak kejahatan.	Menghasilkan perbandingan kinerja dari variasi algoritme <i>fp-tree</i> seperti <i>FP-Growth</i> , <i>COFI-Tree</i> , dan <i>CT-PRO</i> .
8	Atmajaya (2016)	<i>Association rule mining</i>	<i>Apriori</i>	data kejahatan	Menghasilkan aturan yang divisualisasikan dalam bentuk <i>graph</i> .

BAB III LANDASAN TEORI

3.1 *Data Mining*

Perkembangan data yang semakin pesat saat ini disebabkan oleh banyaknya organisasi maupun individu mengumpulkan data secara terus menerus. Hampir semua data tersebut dimasukkan dengan menggunakan aplikasi komputer. Hal tersebut mengakibatkan terjadinya tumpukan data dan munculnya istilah *data rich but information poor*. Karena kondisi tersebut maka dibutuhkan sebuah pengolahan untuk memperoleh informasi atau pola didalamnya melalui suatu analisis. Teknik yang dapat digunakan untuk melakukan analisis dari tumpukan data tersebut adalah *data mining*. Menurut Han, dkk (2012) *data mining* adalah proses menemukan pola yang menarik serta pengetahuan dari sebuah kumpulan data yang besar. Dalam menemukan pola atau pengetahuan digunakan sebuah teknik statistik, matematika, kecerdasan buatan, dan *machine learning* guna mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari data yang besar. Data yang digunakan bisa bersumber dari mana saja selama data tersebut mengandung arti untuk tujuan yang ingin dicapai misalnya saja data yang bersumber dari *database, warehouse, web*, atau *information repositories*.

Menurut Han, dkk (2012) *data mining* dalam menghasilkan pengetahuan harus melalui tahapan proses sebagai berikut:

1) *Data Cleaning and Integration*

Sebelum proses data mining dilakukan, diperlukan proses *data cleaning* dengan maksud menghapus *noise*, membuang duplikasi data, memeriksa data yang inkonsisten dan memperbaiki kesalahan pada data. Proses *enrichment* dan integrasi juga dapat dilakukan untuk memperkaya data yang sudah ada dengan data atau informasi lain yang relevan.

2) *Data selection and Transformation*

Pada tahap *data selection* yang dilakukan adalah melakukan himpunan data, pemilihan data atau memfokuskan subset *variable* atau contoh data yang kemudian data tersebut disimpan dalam suatu berkas terpisah dari basis data operasional. Setelah *data selection* dilakukan tahap selanjutnya yaitu *transformation*, pada tahap

ini dimulai dengan pencarian fitur-fitur yang berguna untuk mempresentasikan data, yang kemudian digunakan pada data yang telah dipilih sehingga data sesuai untuk proses data mining. Proses ini merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

3) *Data Mining*

Tahap *data mining* yakni proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode atau algoritme tertentu. Teknik, metode atau algoritme dalam data mining sangat bervariasi.

4) *Evaluation and Presentation*

Evaluation and Presentation merupakan tahap akhir berupa pengidentifikasian pola yang dihasilkan dari hasil *data mining*. Pola informasi yang dihasilkan dari proses data mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan.

Data mining memfokuskan kajian untuk menemukan pola serta pengetahuan dalam data. Salah satu teknik *data mining* yang dapat digunakan untuk menemukan pola serta hubungan antar data yang ada adalah *association rule mining*. Selain *association rule mining* beberapa fungsi penting yang dapat dilakukan oleh *data mining* antara lain klasifikasi, klustering, estimasi, seleksi variabel, visualisasi dan aturan (Santosa, 2007).

3.2 *Association Rule Mining*

Apabila diperlukan pengetahuan berupa sebuah aturan atau hubungan antar data yang ada pada sebuah basis data maka *association rule mining* digunakan untuk menghasilkan pengetahuan tersebut. *Association rule mining* telah banyak dimanfaatkan untuk melakukan analisis transaksi penjualan atau sering dikenal sebagai *market basket analysis*. Selain itu *association rule mining* juga dapat digunakan pada studi kasus yang lain. *Association rule mining* diperkenalkan pertama kali oleh Agrawal, dkk (1993) sebagai sebuah teknik di dalam *data mining* dengan tujuan untuk melihat hubungan dari *item-item* transaksi di dalam sebuah *database*. Pada saat ini teknik untuk menemukan pola seperti ini sangat penting dalam proses *Knowledge Discovery and Data Mining* (KDD). Bertambahnya data secara terus menerus mendorong banyak pihak menginginkan untuk melakukan analisis terhadap pola pada data yang mereka simpan di dalam sebuah basis data.

Pengetahuan yang diperoleh di dalam himpunan data relasional dapat membantu dalam pengambilan keputusan.

Didalam pencarian pola hubungan menggunakan *association rule mining* terdapat istilah yaitu *itemset*. *Itemset* adalah satu set *item* (Han, dkk., 2012) atau bisa diartikan sebagai himpunan yang beranggotakan satu atau lebih *item*. Suatu *itemset* yang beranggotakan k buah *item* dapat ditulis atau disebut k -*itemset* (Han, dkk., 2012). Sebagai contoh: $\{computer, antivirus_software\}$ merupakan 2-*itemset* karena terdiri dari 2 *item*.

Misalkan terdapat *dataset* $I = \{A, B, \dots, Z\}$ dimana A adalah satu *itemset*. *Itemset* dapat berisi satu *item* atau lebih. *Association rule* mempunyai bentuk sebagaimana ditunjukkan oleh Persamaan (3.1).

$$A \rightarrow B \quad (3.1)$$

Association rule merupakan implikasi bentuk dari $A \rightarrow B$, dimana $A \subset I$, $B \subset I$, $A \neq \emptyset$, $B \neq \emptyset$ dan $A \cap B = \emptyset$. Aturan $A \rightarrow B$ merupakan bagian dari *dataset* I dengan *support* s , dimana s adalah ratio antara banyaknya transaksi dalam I yang mengandung $(A \cup B)$ dengan banyaknya transaksi. Notasi $(A \cup B)$ adalah *union* atau gabungan *itemset* A dan B bukan menunjukkan *itemset* A atau B . Misal *itemset* $A = \{computer\}$, $B = \{antivirus_software\}$, maka $(A \cup B) = \{computer, antivirus_software\}$. Besaran nilai *support* digunakan untuk melihat seberapa besar atau sering *itemset* A dan B muncul di dalam *dataset*. Untuk mendapatkan nilai *support* dari $A \rightarrow B$ dapat dilakukan dengan formula yang diperlihatkan oleh Persamaan (3.2).

$$support(A \rightarrow B) = \frac{\text{banyaknya transaksi yang mengandung itemset } A \text{ dan } B}{\text{banyaknya transaksi}} \quad (3.2)$$

Aturan $A \rightarrow B$ memiliki *confidence* c , dimana c adalah rasio antara banyaknya transaksi yang mengandung *itemset* A dan B dengan banyaknya transaksi yang mengandung *itemset* A . Banyaknya transaksi yang mengandung *itemset* A dan B dikenal sebagai *support count*. Besaran nilai *confidence* digunakan untuk mengukur kekuatan aturan. *Confidence* untuk $A \rightarrow B$ dapat dipresentasikan sebagai probabiliti kondisional, $P(B|A)$. Untuk mendapatkan *confidence* $A \rightarrow B$ digunakan sebuah formula yang diperlihatkan pada Persamaan (3.3).

$$confidence(A \rightarrow B) = P(B|A) = \frac{support(A \cup B)}{support(A)} = \frac{support_count(A \cup B)}{support_count(A)} \quad (3.3)$$