**MANAJEMEN TEKNOLOGI INFORMASI**

**Analisis Penerapan Knowledge Discovery in Database (KDD)**

**Terhadap Kecelakaan Lalu Lintas**



**Oleh :**

**Siti Larista Octaria**

**09031181621128**

**Dosen Pengampu : Deris Setiawan, M.T.,Ph.D**

**PROGRAM STUDI SISTEM INFORMASI**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS SRIWIJAYA**

**2018**

**Knowledge Discovery in Databases (KDD)**

*Knowledge discovery in databases* (KDD) adalah keseluruhan proses non-trivial untuk mencari dan mengidentifikasi pola (pattern) dalam data, dimana pola yang ditemukan bersifat sah, baru, dapat bermanfaat dan dapat dimengerti. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interprestasi dan visualisasi dari pola-pola sejumlah kumpulan data.

*Knowledge Discovery in Databases* merupakan sekumpulan proses untuk menemukan pengetahuan yang bermanfaat dari data. KDD terdiri dari serangkaian langkah perubahan, termasuk data preprocessing dan juga post processing. Data propecessing merupakan langkah untuk mengubah data mentah menjadi format yang sesuai untuk tahap analisis berikutnya. Selain itu data preprocessing juga digunakan untuk membantu dalam pengenalan atribut dan data segmen yang relevan dengan task Data Mining.

Istilah *Data Mining* dan *Knowledge Discovery in Databases* (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah Data Mining. Proses Knowledge Discovery in Databases (KDD) secara garis besar dapat dijelaskan sebagai berikut :



**Tahapan Knowledge Discovery in Databases (KDD) :**

1. ***Data Selection***

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam knowledge data discovery dimulai. Data hasil seleksi yang akan digunakan untuk proses Data Mining, disimpan dalam suatu berkas, terpisah dari basis data operasional.

1. ***Pre-processing atau Cleaning***

Sebelum proses Data Mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus knowledge data discovery. Proses cleaning mencakup antaralain membuang duplikasi data, memeriksa data yang inkosisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan prosesenrichment, yaitu proses memperkaya data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan KDD, seperti data atau informasi eksternal.

1. ***Transformation***

Coding adalah proses tranformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses Data Mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

1. ***Data Mining***

Data Mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam Data Mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses KDD secara keseluruhan.

1. ***Interpretation atau Evaluation***

Pola informasi yang dihasilkan dari proses Data Mining perlu ditampilkan dalam bentuk yang mudah dimengerti oleh pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang disebut interpretation. Tahap ini mencakup pemeriksaan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada sebelumnya.

**ANALISIS PENERAPAN KNOWLEDGE DISCOVERY IN DATABASE (KDD) TERHADAP KECELAKAAN LALU LINTAS**

1. **Analisa Karakteristik Kecelakaan Lalu Lintas di Jalan Ahmad Yani Surabaya melalui Pendekatan Knowledge Discovery in Database**

Analisa karakteristik kecelakaan merupakan tahapan audit keselamatan jalan sebagai bagian manajemen operasional aset Jl. Ahmad Yani Surabaya meliputi analisa situasi, identifikasi *blackspot*, faktor pemicu dan karakteristik kecelakaan. Metode yang digunakan adalah *Knowledge Discovery in Databases* dengan *data mining* didukung aplikasi SPSS dan RapidMiner.

Studi dilakukan dengan pendekatan kuantitatif melalui proses *data mining* dan *plotting*, yang didukung analisa kualitatif untuk pendeskripsian fenomena masalah. Metode *data mining* menggunakan algoritma statistik seperti *trendline*, *cross-tab*, *scattering*, *clustering*, dsb. Sedangkan *data plotting* menggunakan stritmap jalan untuk pemetaan lokasi *blackspot*.







Dengan melakukan *plotting* data (atribut jalur dan titik kejadian) terhadap model stritmap, dapat dikenali lokasi rawan laka (*blackspot*) di Jl. Ahmad Yani. Selain itu, melalui *text mining* pada data yang sama, didapatkan beberapa katakunci mewakili lokasi yang menjadi modus kejadian laka lantas.

Jalan di sekitar Royal Plaza, pertigaan Margorejo dan RSAL menjadi lokasi paling rawan laka lantas. Beberapa faktor pemicu kerawanan pada *blackspot* tersebut antara lain :

1. banyaknya aktivitas keluar-masuk kendaraan dan pejalan kaki (penyeberang jalan) di sekitar pusat kegiatan (mall dan rumah sakit).
2. terdapat percabangan jalur jalan sehingga terjadi pertemuan arah pergerakan kendaraan yang berbeda (termasuk variasi pada kecepatan laju kendaraan).
3. kendaraan yang berpindah lajur secara mendadak atau tanpa sinyal (biasanya untuk berbelok arah, mengakses lokasi tujuan atau pindah jalur jalan).

**Karakteristik Kecelakaan**

1. Karakteristik Kecelakaan dalam Parameter Waktu

Pengamatan data dalam atribut hari tidak menunjukkan adanya nilai signifikan karena prosentase kejadian tidak berbeda jauh dan dalam kisaran 10-16 persen. Kejadian kecelakaan paling banyak terjadi pada hari Rabu (15,9%). Namun, bila atribut dikaitkan dengan jam kejadian, maka waktu paling berpotensi kecelakaan adalah hari Senin pukul 20:00-21:00 (malam). Bila dicermati, kejadian kecelakaan tidak banyak terjadi pada jam-jam sibuk (*peak hour*). Kemacetan tidak memberikan peningkatan terhadap potensi laka lantas. Kecelakaan terdominasi pada jam-jam lalu lintas lenggang, yakni pukul 20:00-22:00 (9,5%) dan 11:00-12:00 (7,7%). Pada waktu tersebut kondisi arus bebas tercapai dan menyebabkab pengendara melaju dengan kecepatan tinggi, khususnya pada *frontage road*.

1. Karakteristik Kecelakaan dalam Parameter Pihak Terlibat dan Fatalitas

Dari seluruh data kecelakaan yang ada, fatalitas kecelakaan masih dominan pada skala rendah (75,9%), dimana korban kecelakaan sebagian besar hanya luka ringan (79,38%). Namun dengan masih adanya korban jiwa mencapai angka 10%, maka sebaiknya kejadian kecelakaan fatal perlu diminimalkan. Analisa korelasi pada atribut fatalitas kecelakaan terhadap jumlah korban menunjukkan nilai positif dan hasil yang signifikan, yakni *Pearson* =.298 (sig.0.00); *Kendall’s tau* =.241 (sig.000); dan *Sparman’s rho*=.256 (sig.0.00).

1. Karakteristik Kecelakaan dalam Parameter Biaya Kecelakaan

Analisa menunjukkan bahwa biaya korban dan kerugian materiil terus meningkat dengan signifikan, meskipun pada tahun 2016 kejadian kecelakaan menurun frekuensinya. Prosentase peningkatan kerugian materiil lebih besar dibandingkan biaya korban kecelakaan mencapai 466,49% .

Perhitungan kerugian kecelakaan di Jl. Ahmad Yani selama periode data menunjukkan nilai yang sangat besar, yakni mencapai nilai Rp. 10,78 Triliun. Kondisi ini mendasari pentingnya dilakukan audit keselamatan jalan dan tindakan pencegahan secara komprehensif (aspek teknis dan non teknis). Selain itu, adanya korban jiwa sebenarnya merupakan kerugian yang jauh tidak ternilai, mengingat nyawa manusia bukanlah barang yang dapat diperbaiki atau pun digantikan.

1. **Analisa Asosiatif Data Mining Untuk Mengetahui Pola Kecelakaan Lalu Lintas**

Data kecelakaan lalu lintas dapat diolah menjadi informasi yang penting bagi Kepolisian. Informasi diperoleh dengan cara menganalisis data kecelakaan lalu lintas untuk mengetahui apakah ada hubungan antara terjadinya kecelakaan untuk merek kendaraan dan jenis kendaraan. Penelitian ini menerapkan metode data mining untuk memproses kecelakaan lalu lintas data dengan menggunakan teknik data mining yang disebut Metode Apriori. Penelitian ini mengambil lokasi ruas jalan Slamet Riyadi-PB.Soedirman-Sultan Agung-Gajah Mada-Hayam Wuruk di Kabupaten Jember.

Metode apriori digunakan untuk memperoleh kaidah asosiasi yang menggambarkan hubungan antar item pada database transaksional. Apriori Metode digunakan untuk mengidentifikasi pola kecelakaan berdasarkan merek dan jenis kendaraan. Analisa data mining terhadap data kendaraan bermotor dan data kecelakaan diharapkan dapat memenuhi beberapa kebutuhan yaitu mengemukakan pola kecelakaan kendaraan bermotor berdasarkan pola merek, jenis kendaraan, dan warna.

 **Pola Kecelakaan**

1. Pola kecelakaan untuk wilayah DIY berdasarkan merek





Pola kecelakaan berdasarkan merek hanya memperhatikan hubungan antara merek dengan kejadian kecelakaan. Tabel ini memperlihatkan jumlah kejadian kecelakaan bagi merek-merek kendaraan bermotor menggunakan minimum support 10%.

1. Pola Kecelakaan untuk wilayah DIY berdasarkan tipe kendaraan



Pola kecelakaan berdasarkan tipe hanya memperhatikan hubungan antara tipe kendaraan dengan kejadian kecelakaan. Jumlah kejadian kecelakaan berdasarkan tipe kendaraan menggunakan minimum support 5%.

Setelah proses penelitian dilakukan diperoleh beberapa kesimpulan tentang hasil data mining terhadap data penjualan kendaraan bermotor dan data kecelakaan sebagai berikut :

1. Pola kecelakaan telah diketahui. Kecelakaan sering terjadi untuk kendaraan dengan merek Suzuki, disusul Yamaha dan Kawasaki.
2. Jika digunakan kriteria jenis kendaraan, kecelakaan sering terjadi pada sepeda motor, disusul dengan sedan dan kemudian minibus.
3. Pola kecelakaan berdasarkan warna kendaraan tidak dapat dilakukan karena data

warna kendaraan yang terlibat kecelakaan tidak dicatat dalam laporan polisi.

1. **Implementasi Apriori Pada Data Kecelakaan Lalu Lintas Dalam Pencarian Relasi Antar Variabel Pelaku Laka**

Penanggulangan kecelakaan lalu lintas dapat dilakukan dengan melakukan langkah – langkah pendekatan seperti sosialisasi dan pengawasan terhadap rambu – rambu serta trafik pada jalan raya. Dengan mengamati variabel – variabel yang berkaitan dari para pelaku kecelakaan lalu lintas maka dapat diperoleh kriteria dari para pelaku kecelakaan lalu lintas yang dominan dari relasi variabel – variabel pelaku LAKA seperti pendidikan, umur dan variabel lainnya.

Algoritma apriori menjadi pelopor dalam algoritma analisis asosiasi untuk menemukan kandidat itemset frekuen dan pembangkitan aturan asosiasi untuk menemukan kandidiat itemset frekuen dan pembangkitan aturan asosiasi dapat dibentuk.

Data pelaku kecelakaan lalu lintas yang digunakan dapat dilihat pada tabel yang mana data yang digunakan adalah data pelaku kecelakaan lalu lintas berdasarkan usia, profesi dan pendidikan. Berikut sample data dari pelaku kecelakaan lalu lintas pada tahun 2011.







Pelaku LAKA dengan Profesi TNI yang berjumlah Rendah maka akan berdampak pada jumlah Pelaku LAKA Dengan Profesi Lain - Lain yang Rendah, Sehingga dapat diperoleh kaitan variabel antara profesi dan pendidikan khususnya pada profesi TNI dan pendidikan Sekolah Dasar dalam bentuk keterkaitan jumlah pelaku yang rendah.

Pelaku LAKA denga Profesi Pedagang yang berjumlah Rendah maka akan berdampak pada jumlah Pelaku LAKA Dengan Profesi Lain - Lain yang Rendah dapat diambil kesimpulan profesi Pedagang juga berkaitan dengan pendidikan Lain – Lain dalam bentuk keterkaitan jumlah pelaku yang rendah.

Berdasarkan hasil yang diperoleh dapat dilihat kaitan antar variabel yang dihasilkan cenderung menghasilkan variabel dengan jumlah penyebab LAKA yang rendah sehingga dapat dijadikan informasi bahwa variabel – variabel yang diperoleh diatas seperti Profesi TNI dan Pedagang serta Pendidikan Sekolah Dasar dan Lain – Lain tidak menjadi variabel utama dalam tingginya angka LAKA yang terjadi di Kota Medan untuk tahun 2011.

Berdasarkan proses Apriori maka dapat diperoleh berdasarkan data LAKA sementara pada tahun 2016 maka beberapa variabel pelaku LAKA yang berkaitan dapat dijabarkan sebagai berikut : Profesi Mahasiswa, Pengemudi, Pedagang, Petani/Buruh dan profesi lain-lain salin berkaitan dimana jika angka variabel LAKA Mahasiswa rendah, maka variabel pengemudi, pedagang, petani/buruh dan profesi lain-lain juga pasti akan rendah sehingga terdapat fenomena keterkaitan rendahnya angka pada variabel mahasiswa dan variabel profesi lainnya seperti pengemudi, pedagang, petani/buruh dan profesi lain-lain.

1. **Penentuan Pola Hubungan Kecelakaan Lalu Lintas Menggunakan Metode *Association Rules D*engan Algoritma *Apriori***

Pada proses perhitungan *Association rules* sebelumnya ditentungan dulu data set yang dipilih. Setelah memilih dataset, langkah selanjutnya pengguna menentukan nilai minimum *support* dan minimum *confidence* yang akan digunakan dalam proses. Nilai minimum *support* dan minimum *confidence* yang dimasukkan adalah antara 0 –100%. Penelitian yang dilakukan oleh penulis menggunakan nilai minimal *support* 0,2(20%), dan minimal *confidence* 0,9 (90%).

Algoritma *apriori* adalah algoritma yang sangat mendasar dalam mencari *frequent itemset* dari transaksi *database* berukuran besar. Nama algoritma *apriori* diambil dari kenyataan bahwa algoritma ini menggunakan pengetahuan sebelumnya (*prior knowladge*) dari *frequent itemset* untuk proses iterasi berikutnya.

Variabel penelitian adalah suatu yang menjadi objek penelitian atau juga diartikan sebagai faktor-faktor yang berperan dalam peristiwa atau gejala yang akan diteliti. Variabel respon (*decision*) yang digunakan adalah tingkat keparahan korban kecelakaan lalu lintas yang terdiri dari tiga kategori, yaitu Luka Ringan,Luka Berat, Meninggal Dunia. Ditunjukkan dari tabelJumlah dan Persentase Tingkat Luka Korban Kecelakaan Lalu Lintas. Dari 702 kecelakaan yang terjadi diketahui 625 mengalami Luka Ringan atau sekitar 89% dari total kecelakaan, 30 Luka Berat atau sekitar 4% dari total kecelakaan, dan yang terakhir yaitu 47 Meninggal Dunia atau sekitar 7% dari total kecelakaan.



*Association rules* merupakan salah satu teknik dalam data mining untuk menentukan pola hubungan “jika-maka” antara suatu kombinasi *item Association rules* akan dicari menggunakan algoritma *Apriori,* dengan batasan minimal *support* yang peneliti tentukan yaitu 0,1 dan batasan minimal *confidence* sebesar 0,7 dengan 4 iterasi.

Aturan asosiasi selanjutnya yaitu dengan menggunakan nilai *support* yang sama 0,1 dan nilai*confidence* 0,7 dengan itesasi sebanyak 5 kali( *Large Itemset).* Penelitian ini hanya akan dipilih ataudipertahankan aturan asosiasi yang kuat tingkatkepercayaannya yaitu dengan nilai minimum *confidence* 0,9=905%, minimum *support=* 0,2=20%. Atauran asosiasi yang memenuhisyarat tersebut dengan 3 iterasi menghasilkan 5 aturan asosiasi dengan nilai *support* (tingkat dominasi *itemset)=*0,2=20%, nilai *confidence(*tingkat kepercayaan)=0,9=90%, sedangkan dengan nilai *support* yang sama dilakukan iterasi sebanyak 4 kali (*Large 4 itemset) yaitu* nilai *support* (tingkat dominasi *itemset)=*0,2=20%, nilai *confidence(*tingkat kepercayaan)=0,9=90% menghasilkan 1 aturan asosiasi dimana dengan Sim yang ada, jenis kelamin laki-laki, dan profesi swasta akan menghasilkan tingkat luka yaitu Luka Ringan.

Dengan begitu diperoleh beberapa informasi diantaranya.,Jenis kelamin laki-laki memiliki nilai support lebih tinggi yang mempengaruhi tingkat kecelakaan lalu lintas dan mengakibatkan tingkat luka yaitu Luka Ringan, hal ini ditandai dengan ukuran lingkaran yang lebih besar dibandingkan ukuran lingkaran lainnya. Serta Profesi swasta,mahasiswa/pelajar, dan memiliki sim memiliki *lift rasio* yang lebih tinggi digambarkan dengan warna lingkaran yang lebih hitam dibandingkan dengan aturan pada usia muda.

1. **Pemanfaatan Algoritma K-Means Untuk Pemetaan Hasil Klasterisasi Data Kecelakaan Lalu Lintas**

Pencarian pengetahuan dalam data, dikenal juga sebagai *Knowledge discovery in Database* (KDD), didefinisikan sebagai ekstraksi data yang memiliki potensi informasi berharga yang implisit dan tidak dikenali sebelumnya.

Tahap pra-prosesing adalah pemilihan kriteria (*selection*) dan normalisasi data *(transformation*) sehingga data siap untuk diklasterisasi. Untuk kebutuhan komputasi, informasi yang bersifat deskriptif (nominal) perlu diubah dalam bentuk tipe data ordinal, interval, atau rasio. Dalam hal ini berlaku proses simplifikasi dan generalisasi data untuk mendapatkan informasi yang dapat diolah lanjut. Setelah melalui tahapan simplifikasi dan generalisasi, langkah selanjutnya adalah transformasi data, yaitu proses data dinormalisasi ke dalam rentang nilai

0-1. Normalisasi data sangat dibutuhkan sebelum proses data mining, agar tidak ada parameter yang mendominasi dalam perhitungan jarak antar data.

Dalam penelitian ini digunakan teknik klasterisasi dengan algoritma K-Means yang bertujuan mengelompokkan data berdasarkan kemiripan karakteristik yang dimiliki oleh data tersebut, Adapun model klasterisasi data sesuai dengan algoritma K-Means untuk membagi data.

Setelah proses klasterisasi telah selesai dilaksanakan maka masuk pada tahap ketiga pos-prosesing data berupa analisis dan visualisasi hasil. Dalam penelitian ini, dibutuhkan peran seorang pakar yang memahami dengan baik karakteristik data kecelakaan lalu lintas untuk membandingkan hasil klastering dan menentukan status dari objek-objek jalan di Kabupaten Sleman, DIY.

Sistem klasterisasi data kecelakaan lalu lintas dengan metode K-Means dapat digunakan untuk mengelompokkan obyek jalan berdasarkan kesamaan karakteristik pada jumlah korban, jumlah kendaraan yang terlibat, dan jumlah kecelakaan yang terjadi dalam suatu rentang waktu tertentu dan Sistem klasterisasi data kecelakaan lalu lintas membutuhkan peran seorang pakar lalu lintas untuk menganalisis hasil klastering dan menentukan klasifikasi berupa status tingkat kerawanan jalan raya.