

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Padatnya jumlah penduduk dapat menimbulkan beberapa permasalahan, salah satunya adalah permasalahan transportasi yang merupakan cerminan dari daerah yang padat penduduk. Lonjakan kepadatan penduduk sering terjadi di kota-kota metropolitan yang menjadi pusat populasi suatu provinsi atau daerah. Kota Semarang adalah salah satunya, dimana ibu kota provinsi Jawa Tengah yang juga termasuk dalam 5 kota terbesar di Indonesia ini memiliki jumlah kepadatan penduduk sejumlah 1,680,417 jiwa (Dinas Kependudukan Dan Catatan Sipil Kota Semarang / Juni 2020). Bertambahnya penduduk juga berdampak pada kebutuhan transportasi pribadi yang semakin tinggi dikarenakan pertimbangan kecepatan sampai ditujuan lebih cepat dibandingkan dengan transportasi umum. Sehingga memunculkan salah satu permasalahan transportasi yang sering terjadi selain kemacetan, yaitu: kecelakaan lalu lintas. Dalam laporan tahunan Satlantas Kota Semarang dari tahun 2017, 2018, dan 2019 berurutan telah terjadi sebanyak 936 kasus, 1.008 kasus dan 1.365 kasus. Kemudian pada tahun 2020 dari hasil rekap laporan tahunan laka lantasi Satlantas Polrestabes Semarang menunjukkan jumlah 939 kasus, meliputi 156 korban meninggal dunia, dan 928 korban luka ringan, dengan kerugian materil mencapai Rp. 387.950.000.

Diharapkan, dari pengolahan data mentah yang sudah disebutkan diatas, dapat mempermudah pihak-pihak berwenang dalam pengelompokan kriteria-kriteria yang ada dengan upaya untuk memperkecil angka kecelakaan lalu lintas di masa mendatang. Pengolahan yang dimaksud penulis disini ialah pengolahan data mining menggunakan pengelompokan *clustering*.

Data Mining adalah suatu istilah yang digunakan untuk menggali informasi yang berharga dan berguna di dalam database. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk menginteraksi mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terikat dari berbagai database besar (Katrina et al., 2019). Clustering merupakan salah satu teknik dari salah satu fungsionalitas data mining, algoritma Clustering merupakan algoritma pengelompokan sejumlah data menjadi kelompok-kelompok data tertentu (Cluster). Setiap Cluster memiliki Centroid yang merupakan suatu besaran yang dihitung dari rata-rata nilai tiap items dari suatu Cluster dan juga memiliki medoid yang merupakan item yang letaknya paling tengah. Sementara jarak antar Cluster didefinisikan dengan menggunakan beberapa metode-metode untuk menentukan Cluster mana yang berdekatan (Sibuea & Safta, 2017). *Clustering* bertujuan untuk mengelompokkan data yang memiliki kemiripan antar satu dengan yang lainnya kedalam suatu klaster atau kelompok sehingga data tersebut memiliki tingkat kemiripan yang maksimal dalam suatu klaster dan mempunyai kemiripan minimal antar klaster. Dengan kata lain *clustering* ditujukan untuk menemukan kelompok objek sehingga

objek-objek didalam kelompok akan sama satu sama lainnya dan berbeda dari objek-objek didalam kelompok lain.

Pada penelitian dengan judul “**PERBANDINGAN CLUSTERING PADA SEGMENTASI KECELAKAAN LALU LINTAS DI KOTA SEMARANG MENGGUNAKAN ALGORITMA K-MEANS DAN ALGORITMA K-MEDOIDS**” ini penulis melakukan *clustering* dengan menggunakan dua algoritma clustering yaitu algoritma K-Means dan algoritma K-Medoids. Data yang digunakan pada penelitian kali ini ialah data kecelakaan lalu lintas di kota Semarang selama dua belas bulan terakhir yang didapat dari rekap laporan tahunan Satlantas Polrestabes Semarang.

1.2 PERUMUSAN MASALAH

Berdasarkan penjelasan latar belakang diatas, maka dapat disimpulkan untuk rumusan masalahnya adalah bagaimana membentuk segmentasi korban kecelakaan lalu lintas berdasarkan data yang ada dengan beberapa kriteria seperti: terjadinya kecelakaan (waktu & hari), karakter korban kecelakaan (pekerjaan, jenis kelamin, dan umur), jenis kecelakaan dan kendaraan yang terlibat kecelakaan dengan menggunakan perbandingan hasil cluster menggunakan algoritma K-Means dan algoritma K-Medoids.

1.3 BATASAN MASALAH

Batasan-batasan dalam penelitian ini ialah seperti berikut:

- 1.3.1 Menggunakan data kecelakaan lalu lintas yang didapat dari hasil rekap laporan tahunan Satlantas Polrestabes Semarang.
- 1.3.2 Menggunakan perbandingan hasil cluster menggunakan dua algoritma antara algoritma K-Means dengan algoritma K-Medoids.
- 1.3.3 Menggunakan Ms. Excel untuk perhitungan manual dan Bahasa pemrograman R dengan software RStudio Clouds untuk perhitungan komputasi.

1.4 TUJUAN DAN MANFAAT PENELITIAN

Tujuan dan manfaat pada penelitian kali ini adalah untuk membandingkan hasil cluster algoritma K-Means dengan algoritma K-Medoids menggunakan data kecelakaan lalu lintas di Kota Semarang dan memperoleh hasil segmentasi dari hasil perhitungan data kecelakaan lalu lintas di Kota Semarang menggunakan perbandingan algoritma K-Means dengan algoritma K-Medoids yang dapat digunakan sebagai referensi pihak terkait dalam upaya menurunkan angka kecelakaan di Kota Semarang pada masa mendatang memanfaatkan hasil segmentasi yang sudah dihasilkan.

1.5 METODE PENELITIAN

1.5.1 Objek Penelitian

Pada penelitian kali ini penulis menggunakan data kecelakaan lalu lintas di kota Semarang dalam jangka 12 bulan terakhir yang diperoleh dari Satlantas Polrestabes Kota Semarang.

1.5.2 Metode Pengumpulan Data

1.5.2.1 Data Primer

Data Primer berasal dari data kecelakaan lalu lintas di kota Semarang dalam jangka 12 bulan terakhir yang diperoleh dari Satlantas Polrestabes Kota Semarang.

1.5.2.2 Data Sekunder

Penyatuan data didapat dari study literature, jurnal, buku, referensi yang berasal dari internet, dan juga wawancara yang bisa membantu penulis dalam menyelesaikan penelitian ini.

1.5.2.3 Studi Pustaka

Metode pengumpulan data yang berasal dari berbagai macam literatur yang masih relevan dengan objek penelitian.

1.5.3 Metode Analisis Data

Metode analisis data yang akan dipakai ialah CRIPS-DM. CRIPS-DM merupakan keseluruhan tahapan penelitian yang digunakan untuk melakukan proses analisis dan prediksi dalam memecahkan suatu permasalahan (Indra Purnama, Ragil Saputra, Adi Wibowo 2015). Berikut adalah tahapan dalam CRIPS-DM, yaitu:

a. *Business Understanding* / pemahaman bisnis.

Pada fase ini menentukan tujuan peneliti. Tujuan dari penelitian ini adalah untuk menerapkan metode data mining menggunakan Algoritma *K-Means* dan Algoritma *K-Medoids* menggunakan data kecelakaan lalu lintas di kota Semarang dalam jangka 12 bulan terakhir yang diperoleh dari Satlantas Polres Kota Semarang selanjutnya menjelaskan tujuan yang diharapkan sehingga mendapatkan strategi awal untuk mencapai tujuan.

b. *Data Understanding* / pemahaman data.

Pada fase ini dilakukan pengumpulan data awal, selanjutnya menganalisis data dan memahami isi seluruh data jika data ada yang kosong maka harus dilakukan pembersihan data. setelah data sudah di pahami secara menyeluruh lalu dilakukan proses pengolahan data.

c. *Data preparation* / persiapan data.

Fase ini adalah tahapan untuk menyiapkan data yang sudah dipahami Dan melakukan proses pembersihan data, pemilihan atribut, transformasi data dan standarisasi data. Semua data yang sudah siap akan siap untuk diproses.

d. *Modeling* / Pemodelan.

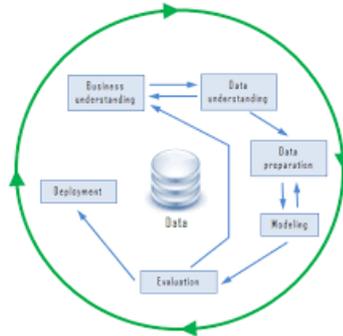
Fase ini adalah fase menentukan teknik permodelan yang sesuai dengan penelitian dan dapat menggunakan beberapa teknik pemodelan untuk menyelesaikan permasalahan. Pada penelitian ini penulis menggunakan *Clustering* dan algoritma *K – Means* dan algoritma *K – Medoids*.

e. *Evaluation* /Evaluasi.

Fase evaluasi adalah fase dimana dilakukan evaluasi terhadap teknik pemodelan yang digunakan, menentukan apakah Teknik pemodelan yang dipilih sudah sesuai dengan tujuan awal dan melihat apakah ada permasalahan penting dalam penelitian yang tidak terselesaikan dengan baik dan mengambil kesimpulan dari hasil yang sudah terselesaikan. Dalam penelitian ini tidak dilakukan tahap evaluasi.

f. *Deployment* / penyebaran.

Pada fase ini adalah penyusunan laporan dan mempresentasikan hasil penelitian menggunakan teknik pemodelan yang sudah digunakan selanjutnya menjelaskan kepada pihak yang bersangkutan. Pada penelitian ini tidak dilakukan tahap penyebaran.



Gambar 1.1 Model tahapan proses CRISP-DM

1.6 SISTEMATIKA PENULISAN

Demi melancarkan pemahaman mengenai laporan riset ini, pengkajian yang selanjutnya disampaikan akan terbagi pada sebagian bab melalui sistematika penyampaian seperti berikut ini:

BAB I PENDAHULUAN

Mengandung hasil penjelasan mengenai latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Berisi tentang pembahasan penelitian yang ada sebelumnya yang berhubungan dengan masalah penelitian yang sedang dilakukan dan berisi materi atau metode yang membantu peneliti menyelesaikan permasalahan .

BAB III METOTOLOGI PENELITIAN

Membahas tentang metode penelitian yang digunakan dan proses penelitian yang dilakukan.

BAB IV HASIL PENELITIAN

Berisi penjelasan hasil penelitian mengenai penelitian yang dilakukan.

BAB V PEMBAHASAN

Berisi penjelasan analisa dan penjabaran penelitian mengenai penelitian yang dilakukan.

BAB VI KESIMPULAN DAN SARAN

Tentang kesimpulan penelitian dan terdapat saran dari peneliti mengenai penelitian yang dibuat dan berisi solusi dari peneliti yang dapat dilanjutkan oleh penelitian selanjutnya.