

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Kecelakaan lalu lintas menurut UU RI No. 22 Tahun 2009 adalah suatu peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan kerugian harta benda (Wordpress, 2016). Kecelakaan lalu lintas di Indonesia dalam tiga tahun terakhir ini menjadi pembunuh terbesar ketiga setelah penyakit jantung koroner dan tuberculosis berdasarkan penilaian oleh WHO (Hidayati & Hendrati, 2016).

Indonesia merupakan negara berkembang dan berpenghasilan menengah dengan populasi penduduk terbanyak ke empat di dunia. Menurut *The Global Report on Road Safety* tahun 2015, Indonesia menduduki peringkat ketiga se-Asia untuk jumlah kematian terbanyak akibat kecelakaan lalu lintas di bawah Tiongkok dan India dengan total 38.279 kematian. Faktanya, menurut *The Global Report on Road Safety* tahun 2015, jika dilihat dari persentase statistik jumlah populasi, Indonesia menduduki peringkat pertama dengan angka kematian akibat kecelakaan lalu lintas sebesar 0,015% 2017 (Nastiti, 2017).

Provinsi Jawa Tengah merupakan salah satu provinsi yang memiliki jumlah kepadatan penduduk tertinggi di Indonesia, berdasarkan hasil Data Sensus Jumlah Penduduk Tahun 2010-2020 tersebut didapatkan jumlah penduduk sebesar 34.490.835 jiwa yang tersebar di 35 Kabupaten/Kota dengan di dominasi penduduk

terbanyak di Kabupaten Brebes, Kabupaten Cilacap, dan Kota Semarang (Badan Pusat Statistik, 2017). Menurut data *Traffic Accidents, Victims and Loss in Region of Police of Jawa Tengah* untuk 5 POLRES di POLDA Jawa Tengah di urutan berdasarkan kategori korban meninggal dunia terbanyak terhitung dari tanggal 1 Januari s/d 14 Agustus 2018 diperoleh RES Banyumas korban meninggal dunia sebanyak 252 jiwa, RESTABES Semarang korban meninggal dunia sebanyak 196 jiwa, RES Pati korban meninggal dunia sebanyak 156 jiwa, RES Brebes korban meninggal dunia sebanyak 147 jiwa, RES Demak korban meninggal dunia sebanyak 141 jiwa (Badan Pusat Statistik, 2018). Dilihat dari data angka kecelakaan lalu lintas kabupaten Pati menempati urutan pertama berdasarkan banyaknya kejadian laka dari POLDA Jawa Tengah yaitu 857 kejadian terhitung dari tanggal 1 Januari S/D 14 Agustus 2018. Terhitung mulai dari 1 Januari 2017 sampai 29 Maret 2017, jumlah kecelakaan lalu lintas di Pati mencapai 250 kejadian yang menyebabkan korban meninggal dunia sebanyak 46 orang dengan rata-rata usia korban meninggal ada di umur produktif yaitu antara 15-30 tahun (Koran Sindo, 2016). Kepolisian Republik Indonesia memiliki data-data kecelakaan lalu lintas hasil dari pencatatan setiap peristiwa kecelakaan yang terjadi. Data-data tersebut perlu dikelola dalam suatu basis data untuk memudahkan proses penggalian informasi-informasi yang ada didalamnya (Dewi, 2018).

Berdasarkan data *Traffic Accidents, Victims and Loss in Region of Police of Jawa Tengah*, Kabupaten Pati memiliki tingkat kecelakaan lalu lintas yang tinggi, maka diperlukan sebuah penelitian tentang pola tingkat keparahan korban kecelakaan lalu lintas. Beberapa penelitian tentang kecelakaan lalu lintas telah

banyak dibahas oleh beberapa peneliti sebelumnya seperti (Indriani & Indawati, 2006) memodelkan dan mengestimasi tingkat kecelakaan Kota Surabaya dengan menggunakan metode log linier dua dimensi, hasil yang di dapatkan terdapat hubungan antara tingkat keparahan korban kecelakaan lalu lintas dengan jenis kendaraan dan waktu terjadinya kecelakaan dengan musim terjadinya kecelakaan.. Penelitian Selanjutnya (Apriani, 2011) pemodelan tingkat kecelakaan lalu lintas di kota Pekanbaru menggunakan Metode Time Series Autoregressive yang hasilnya mengalami penurunan 26 kejadian setiap bulannya di Tahun 2011. Data mining yang membuat data-data kecelakaan menjadi sumber untuk suatu model yang bisa digunakan untuk memprediksi suatu kejadian. Berdasarkan kebutuhan akan pencarian informasi tentang kecelakaan yang melibatkan beberapa kriteria yang tidak bisa ditentukan sebelumnya, maka penggunaan metode data mining merupakan solusi yang layak untuk diajukan (Yunanto, Hariadi, & Purnomo, 2012).

Data mining (penambangan data) adalah suatu proses untuk menemukan suatu pengetahuan atau informasi yang berguna dari data berskala besar (Santosa, 2007). Klasifikasi merupakan salah satu teknik dalam data mining. Klasifikasi adalah pemrosesan untuk menemukan sebuah model atau fungsi yang menjelaskan dan mencirikan konsep atau kelas data, untuk kepentingan tertentu. Klasifikasi adalah tugas dasar dari analisis data yang berfungsi memberikan label kelas untuk kasus yang dijelaskan oleh satu set atribut. Klasifikasi merupakan teknik penting dalam data mining, teknik ini dapat memprediksi label kelas. Sehingga tujuan dari klasifikasi adalah kebenaran dalam memprediksi sebuah nilai (Sarangi & Jaglan, 2013).

Penelitian sebelumnya mengenai *Naïve Bayes* dan *K-Nearest Neighbor* telah banyak digunakan, diantaranya (Saleh, 2015) dengan judul penelitian Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, dan diperoleh hasil persentase 78.33 %. Selanjutnya (Qodrat, 2017) melakukan Perbandingan Algoritma *Naive Bayes* dan *K-Nearest Neighbor* untuk Sistem Kelayakan Kredit pada Nasabah, hasil yang diperoleh Naive Bayes menghasilkan akurasi 84.12 % dan K-Nearst Neighbor menghasilkan akurasi 83.19 %.

Algoritma *Naïve Bayes* memiliki beberapa keunggulan seperti mudah serta biaya perhitungannya kecil (Wu & Kumar, 2009), memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Luthfi & Kusrini, 2009). Namun, semua atribut pada metode ini memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain (Kusumadewi, 2009).

*K-Nearest Neighbor* (KNN) adalah sebuah metode supervised yang berarti membutuhkan data training untuk mengklasifikasikan objek yang jaraknya paling dekat. Prinsip kerja *K-Nearest Neighbor* adalah mencari jarak terdekat antara data yang akan di evaluasi dengan k tetangga (*neighbor*) dalam data pelatihan (Whidhiasih, 2013), Kelebihan dari K-NN adalah dapat digunakan untuk memecahkan permasalahan multiclass (Reaz & Aburomman, 2016). Namun, K-NN memiliki masalah untuk menemukan tetangga terdekat pada titik query dari dataset yang digunakan (Liaw, Wu, & Leou, 2010).

Berdasarkan uraian diatas dapat dipahami kinerja dari algoritma *Naïve Bayes* dan *K-NN*, dengan mengklasifikasi dan menghitung tingkat akurasi data lalu lintas dari SATLANTAS unit LAKA POLRES PATI tahun 2017, penulis akan melakukan perbandingan terhadap algoritma *Naïve Bayes* dan *K-Nearest Neighbor*.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah yang akan dibahas dalam penelitian ini adalah :

1. Bagaimana karakteristik tingkat keparahan korban kecelakaan lalu lintas di Kabupaten Pati Jawa Tengah Tahun 2017?
2. Bagaimana mengetahui klasifikasi terbaik antara algoritma *Naive Bayes* dan *K-Nearest Neighbor* ?

## 1.3 Tujuan Penelitian

Tujuan yang akan dicapai dalam penelitian ini adalah :

1. Mengetahui karakteristik tingkat keparahan korban kecelakaan lalu lintas di Kabupaten Pati Jawa Tengah Tahun 2017.
2. Memodelkan klasifikasi terbaik antara Algoritma *Naive Bayes* dan *K-Nearest Neighbor* .

## 1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian yang hendak dicapai, maka penelitian ini diharapkan mempunyai manfaat sebagai berikut :

1. Manfaat Teoritis

Secara teoritis hasil penelitian ini diharapkan dapat bermanfaat sebagai berikut:

- a. Memebrikan sumbangan ilmiah dalam ilmu pengetahuan, yaitu klasifikasi tingkat keparahan korban kecelakaan lalu lintas di Kabupaten Pati Jawa

Tengah 2017 menggunakan Algoritma *Naive Bayes* dan *K-Nearest Neighbor*.

- b. Sebagai referensi pada penelitian-penelitian selanjutnya yang berhubungan dengan Algoritma *Naive Bayes* dan *K-Nearest Neighbor*

## 2. Manfaat Praktis

Secara Praktis penelitian ini dapat bermanfaat bagi peneliti agar peneliti mampu menerapkan Algoritma *Naive Bayes* dan *K-Nearest Neighbor* yang sesuai dalam materi yang telah dipelajari, serta peneliti mempunyai pengetahuan dan wawasan Algoritma *Naive Bayes* dan *K-Nearest Neighbor*

### 1.5 Batasan Masalah

Batasan masalah pada penelitian ini adalah :

1. Data yang digunakan adalah data sekunder dari SATLANTAS POLRES Pati tentang laporan kecelakaan lalu lintas pada bulan Agustus Tahun 2017.
2. Variabel atribut yang digunakan pada penelitian ini yaitu jenis kelamin, jenis kecelakaan, peran korban, jenis kendaraan, waktu kejadian, jam kejadian, pekerjaan korban, usia korban, faktor manusia, alat keselamatan dan faktor manusia. Sedangkan label yang digunakan pada penelitian ini yaitu tingkat keparahan korban kecelakaan lalu lintas dengan kategori meninggal dan tidak meninggal.
3. Penelitian ini menggunakan penentuan nilai K dengan bilangan ganjil dan ditentukan nilai K yaitu 3.