

BAB I

PENDAHULUAN

1.1 Latar Belakang

Berdasarkan pasal 93 Peraturan Pemerintah No. 43 Tahun 1993 ayat 1 tentang Prasarana Jalan Raya dan Lalu Lintas, kecelakaan lalu lintas diartikan sebagai suatu peristiwa di jalan raya yang tidak diduga dan tidak disengaja, melibatkan kendaraan dengan atau tanpa pemakai jalan lainnya dan menyebabkan korban manusia atau kerugian harta benda. Korban kecelakaan lalu lintas dapat berupa luka ringan, korban mati dan luka berat yang diperhitungkan paling lama 30 (tiga puluh) hari setelah kecelakaan terjadi. Menurut WHO (2013), kecelakaan lalu lintas adalah kejadian dimana sebuah kendaraan bertabrakan dengan benda lain dan menyebabkan kerusakan. Kecelakaan ini dapat menyebabkan luka-luka, kematian manusia atau binatang. Kecelakaan lalu lintas menelan korban jiwa sekitar 1,2 juta manusia pada tiap tahunnya. Organisasi Kesehatan Dunia (WHO) memprediksikan pada tahun 2030 kecelakaan lalu lintas akan menjadi faktor pembunuh manusia paling besar kelima di dunia. Penelitian yang dilakukan oleh Fitriah dkk., (2012) bahwa Dinas Perhubungan menyatakan kecelakaan lalu lintas menjadi penyebab kematian nomor tiga di Indonesia setelah serangan jantung dan stroke.

Pendataan Kecelakaan Lalu Lintas di Kabupaten Pati menempati urutan pertama se-Jawa Tengah (MitraPost, 2017). Kanit Lakalantas Polres Pati, Ipda Inung Hesti Yugastanto mengatakan bahwa jumlah laka lintas di Kabupaten Pati dari tahun ke tahun mengalami peningkatan. Pada tahun 2017 tercatat telah terjadi

kasus kecelakaan lalu lintas sebanyak 1028 kasus, kemudian di tahun 2018 meningkat lagi menjadi 1121 kasus dan di bulan Desember akhir tahun 2019 jumlah kecelakaan kembali naik menjadi 1498 kasus (LintasPantura, 2019).

Data mining adalah salah satu bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengolahan pola, statistik, database dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar (Larose, 2005). Terdapat banyak teknik pengerjaan data mining untuk menemukan pola atau informasi tersembunyi, diantaranya Klasterisasi (*clustering*), Regresi (*regression*), Asosiasi (*association*) dan Klasifikasi (*classification*) (Amalia, 2018).

Metode klasifikasi merupakan salah satu teknik data mining yang bertujuan untuk memisahkan individu atau objek ke dalam kelas-kelas tertentu. Metode klasifikasi yang baik akan menghasilkan sedikit kesalahan klasifikasi atau akan menghasilkan peluang kesalahan alokasi yang kecil (Johnson dan Wichern, 2007). Penelitian mengenai klasifikasi karakteristik kecelakaan lalu lintas telah dilakukan oleh Syaputra (2018). Hasil penelitiannya mengenai klasifikasi kejadian kecelakaan lalu lintas berdasarkan luka korban dengan *support vector machine* pada studi kasus kecelakaan lalu lintas di Sleman, Yogyakarta Tahun 2016-2017. *Support Vector Machine* adalah suatu teknik relatif baru pada tahun 1995 untuk melakukan prediksi, baik dalam klasifikasi maupun regresi yang sangat populer belakangan ini (Fauzi dkk., 2017). Hasil penelitiannya menggunakan *support vector machine* memperoleh tingkat akurasi sebesar 82,264% dengan tingkat kesalahan klasifikasi sebesar 17,735%. Penelitian lain mengenai kecelakaan lalu

lintas juga dilakukan Wisdayani (2019) mengenai Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Jawa Tengah dengan menggunakan Algoritma *K-Nearest Neighbor*. *K-Nearest Neighbor* adalah suatu algoritma pembelajaran mesin sederhana yang didasarkan pada gagasan bahwa suatu objek yang ‘dekat’ satu sama lain yang memiliki karakteristik yang mirip. Hasil penelitiannya memperoleh tingkat akurasi 88,82% dan tingkat kesalahan sebesar 11,18% (Wisdayani, 2019).

Penelitian sebelumnya telah dilakukan oleh Putranto dkk. (2015) mengenai perbandingan klasifikasi *Decision Tree* dan *Support Vector Machine* untuk penentuan jurusan pada siswa SMA. *Decision tree* adalah model prediksi terhadap suatu keputusan menggunakan struktur hirarki atau pohon. Setiap pohon memiliki cabang, cabang mewakili suatu atribut yang harus dipenuhi untuk menuju cabang selanjutnya hingga berakhir di daun (tidak ada cabang lagi) (Jayanti dkk., 2008). Hasil dari penelitiannya memperoleh *Decision Tree* memperoleh tingkat akurasi yang lebih baik dengan nilai 88,57% dan tingkat kesalahan sebesar 11,43% dibandingkan *Support Vector Machine* memperoleh tingkat akurasi sebesar 87,14% dan tingkat kesalahan sebesar 12,86%. Perbandingan metode *naive bayes* dengan *K-Nearest Neighbor* pada klasifikasi artikel berbahasa Indonesia telah dilakukan oleh Devita (2018). *Naive Bayes* termasuk ke dalam pembelajaran *supervised*, sehingga pada tahapan pembelajaran dibutuhkan data awal berupa data pelatihan untuk mengambil keputusan. Pada tahapan pengklasifikasian akan dihitung nilai probabilitas dari masing-masing label kelas yang ada terhadap masukan yang diberikan. Label kelas yang memiliki nilai probabilitas paling besar

yang akan dijadikan label kelas data masukan tersebut. *Naive bayes* merupakan perhitungan teorema bayes yang paling sederhana, karena dapat mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma naive bayes juga mampu menangani set data yang memiliki banyak atribut (Yoo dkk., 2012). Hasil dari penelitiannya memperoleh tingkat akurasi lebih baik sebesar 70% dibandingkan *K-Nearest Neighbor* dengan tingkat akurasi sebesar 30%.

Penelitian mengenai Algoritma C4.5 menggunakan *Adaboost* untuk peningkatan akurasi dalam meminimalkan resiko kredit telah dilakukan oleh (Nurzahputra dan Muslim, 2017). Algoritma C4.5 merupakan pengembangan algoritma ID3 untuk meningkatkan beberapa karakteristik, yaitu: dapat digunakan untuk menangani atribut kontinu, dapat berurusan dengan *data training* yang nilai atributnya hilang serta dapat digunakan untuk memproses pemangkasan pohon yang dibangun (Mantas dan Abellan, 2014). Metode *Adaboost* merupakan salah satu teknik dari *Ensemble Methods / Boosting Methods* yang digunakan untuk menyeimbangkan dan mengkombinasikan *record – record* dalam sebuah kelas dengan cara memberikan koefisien bobot berdasarkan *performance* hasil training sebelumnya untuk meningkatkan tingkat akurasi dengan meminimalkan *function error* pada klasifikasi (Jusia, 2018). *Adaptive Boosting (Adaboost)* merupakan *ensemble learning* yang sering digunakan pada algoritma *boosting*. *Boosting* bisa dikombinasikan dengan *classifier* algoritma yang lain untuk meningkatkan performa klasifikasi. Hasil dari penelitiannya memperoleh akurasi dari 70,5% ke 74,2% yang menunjukkan bahwa terdapat peningkatan akurasi sebesar 3,7%.

Penelitian mengenai *Naive Bayes* menggunakan *Adaboost* untuk peningkatan akurasi pada prediksi penyakit ginjal kronik telah dilakukan oleh Irawan (2016). Hasil penelitiannya menunjukkan peningkatan akurasi sebesar 0,03 dimana dari akurasi 0,95 menjadi 0,98. Berdasarkan uraian latar belakang di atas, penulis tertarik melakukan sebuah penelitian untuk menyusun skripsi yang berjudul “Perbandingan *Adaboost* C4.5 dan *Adaboost Naive Bayes* pada Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati”.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang dibahas dalam penelitian sebagai berikut:

1. Bagaimana gambaran umum mengenai tingkat keparahan kecelakaan lalu lintas di Kabupaten Pati pada tahun 2019?
2. Bagaimana tingkat akurasi *Adaboost* C4.5 pada Kecelakaan Lalu Lintas di Kabupaten Pati?
3. Bagaimana tingkat akurasi *Adaboost Naive Bayes* pada Kecelakaan Lalu Lintas di Kabupaten Pati?
4. Bagaimana perbandingan tingkat akurasi antara *Adaboost* C4.5 dan *Adaboost Naive Bayes* pada Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian yang hendak dicapai sebagai berikut:

1. Untuk mengetahui gambaran umum mengenai tingkat keparahan kecelakaan lalu lintas di Kabupaten Pati pada tahun 2019.
2. Untuk mengetahui tingkat akurasi *Adaboost C4.5* pada Kecelakaan Lalu Lintas di Kabupaten Pati.
3. Untuk mengetahui tingkat akurasi *Adaboost Naive Bayes* pada Kecelakaan Lalu Lintas di Kabupaten Pati.
4. Untuk mengetahui perbandingan tingkat akurasi antara *Adaboost C4.5* dan *Adaboost Naive Bayes* Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati.

1.4 Manfaat Penelitian

Manfaat dari penelitian ini sebagai berikut:

1. Bagi mahasiswa mampu mengaplikasikan ilmu statistika untuk analisis *Adaboost C4.5* dan *Adaboost Naive Bayes*
2. Bagi perpustakaan Jurusan Statistika Universitas Muhammadiyah Semarang, sebagai referensi yang dapat bermanfaat tentang *Adaboost C4.5* dan *Adaboost Naive Bayes*
3. Bagi pembaca mampu menambah ilmu pengetahuan mengenai *Adaboost C4.5* dan *Adaboost Naive Bayes* pada klasifikasi korban kecelakaan lalu lintas.

1.5 Batasan Penelitian

Dalam penelitian ini, peneliti memberikan batasan bahwa dalam penelitian ini membandingkan tingkat akurasi *Adaboost C4.5* dan *Adaboost Naive Bayes* dengan menggunakan data sekunder yang diambil dari Kepolisian Resor

Kabupaten Pati tentang data kecelakaan lalu lintas di Kabupaten Pati pada Tahun 2019 dengan jumlah data sebanyak 207 data yang diambil dari Tugas Akhir Intan Kurnia Dewi dengan menggunakan *tool Rapidminer*.

