



**PERBANDINGAN ADABOOST C4.5
DAN ADABOOST NAIVE BAYES PADA KLASIFIKASI
KORBAN KECELAKAAN LALU LINTAS DI KABUPATEN PATI**

JURNAL ILMIAH

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Statistika

Oleh

DINDA ROSYADA NURUZZULFA

B2A219019

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS MUHAMMADIYAH SEMARANG
2020**

PERSETUJUAN PEMBIMBING

Skripsi dengan judul “*Perbandingan Adaboost C4.5 dan Adaboost Naive Bayes pada Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati*” yang disusun oleh:

Nama : Dinda Rosyada Nuruzzulfa

NIM : B2A219019

Program Studi : S1 Statistika

Skripsi ini telah disetujui oleh pembimbing untuk diajukan ke sidang panitia ujian Skripsi Program Sarjana.

Pembimbing Utama

Pembimbing Pendamping

Tiani Wahyu Utami, S.Si, M.Si

Prizka Rismawati Arum, S.Si, M.Stat

NIK.28.6.1026.341

NIK.CP.1026.071

PENGESAHAN KELULUSAN

Skripsi dengan judul "**Perbandingan *Adaboost C4.5* dan *Adaboost Naive Bayes* pada Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati**" yang disusun oleh:

Nama : Dinda Rosyada Nuruzzulfa
NIM : B2A219019
Program Studi : S1 Statistika

telah diujikan dalam Sidang Panitia Ujian Skripsi Program Sarjana, Universitas Muhammadiyah Semarang pada tanggal 22 September 2020



Anggota Tim Penguji I

Dr. Rochdi Wasono, M.Si
NIK. 28.6.1026.119

Anggota Tim Penguji II

Tiani Wahyu Utami, M.Si
NIK. 28.6.1026.341

Anggota Tim Penguji III

Prizka Rismawati Arum, S.Si, M.Stat
NIK. CP.1026.071

Mengetahui
Ketua Program Studi

Indah Manfaati Nur, S.Si, M.Si
NIK. 28.6.1026.221

**SURAT PERNYATAAN
PUBLIKASI KARYA ILMIAH**

Nama : Dinda Rosyada Nuruzzulfa
NIM : B2A219019
Fakultas /Jurusan : S1 Statistika
Jenis Penelitian : Skripsi
Judul : Perbandingan *Adaboost C4.5* dan *Adaboost Naive Bayes*
pada Klasifikasi Korban Kecelakaan Lalu Lintas di
Kabupaten Pati
Email : zulfa_dinda24@yahoo.com

Dengan ini menyatakan bahwa saya menyetujui untuk :

1. Memberikan hak bebas royalti kepada perpustakaan Unimus atas penulisan karya ilmiah saya, demi pengembangan ilmu pengetahuan
2. Memberikan hak menyimpan, mengalih mediakan/mengalih formatkan, mengelola dalam bentuk pengakalan data(*database*), mendsitribusikannya, serta menampilkannya dalam bentuk *sofcopy* untuk kepentingan akademis kepada perpustakaan Unimus, tanpa perlu menerima ijin dari saya selama tetap mencantumkan nanam saya sebagai penulis/pencipta.
3. Bersedia dan menjamin untuk menanggung secara pribadi tanpa melibatkan perpustakaan Unimus, dari semua bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini.

Demikian pernyataan ini saya buat dengan sesungguhnya dan semoga dapat digunakan sebagaimana mestinya.

Semarang, 22 September 2020
Yang menyatakan,



Dinda
Dinda Rosyada Nuruzzulfa
NIM. B2A219019

**PERBANDINGAN ADABOOST C4.5 DAN ADABOOST NAIVE BAYES
PADA KLASIFIKASI KORBAN KECELAKAAN LALU LINTAS
DI KABUPATEN PATI**
Dinda Rosyada Nuruzzulfa¹, Tiani Wahyu Utami², PrizkaRismawati Arum³

¹²³Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang
Alamat e-mail : zulfa_dinda24@yahoo.com

ABSTRAK

Kecelakaan lalu lintas adalah kejadian dimana sebuah kendaraan bertabrakan dengan benda lain dan menyebabkan kerusakan. Organisasi Kesehatan Dunia memprediksikan pada tahun 2030 kecelakaan lalu lintas akan menjadi faktor pembunuh manusia paling besar kelima di dunia. Jumlah laka lantas di Kabupaten Pati dari tahun ke tahun mengalami peningkatan. Tujuan penelitian ini adalah membandingkan *Adaboost C4.5* dan *Adaboost Naive Bayes* pada Klasifikasi Kecelakaan Lalu Lintas di Kabupaten Pati. Variabel independen yang digunakan adalah Usia, Jenis Kelamin, Pendidikan, Jenis Pekerjaan, Jenis Kecelakaan, Jam Kejadian, Tanggal Kejadian, Alat Keselamatan, Peran Korban, Jenis Kendaraan, dan Faktor Manusia. Berdasarkan pengujian yang telah dilakukan hasil akurasi dari *Adaboost Naive Bayes* lebih baik dibandingkan *Adaboost C4.5*. Dengan demikian dapat disimpulkan bahwa *Adaboost Naive Bayes* baik digunakan dalam Klasifikasi Korban Kecelakaan Lalu Lintas di Kabupaten Pati.

Kata Kunci : Kecelakaan Lalu Lintas, *Adaboost*, *C4.5*, *Naive Bayes*.

ABSTRACT

A traffic accident is an incident where a vehicle collides with another object and causes damage. The World Health Organization predicts by 2030 traffic accidents will be the fifth-largest human killer factor in the world. The number of laka in Pati Regency from year to year has increased. The purpose of this study was to compare Adaboost C4.5 and Adaboost Naive Bayes on the Traffic Accident Classification in Pati District. Independent variables used are Age, Gender, Education, Job Type, Accident Type, Time of Occurrence, Date of Incident, Safety Tool, Victim Role, Vehicle Type, and Human Factor. Based on the testing that has been done the accuracy results of Adaboost Naive Bayes are better than Adaboost C4.5. Thus it can be concluded that Adaboost Naive Bayes is well used in the Classification of Traffic Accident Victims in Pati District

Keywords: Traffic Accidents, *Adaboost*, *C4.5*, *Naive Bayes*

PENDAHULUAN

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. Menurut WHO (2013), kecelakaan lalu lintas adalah kejadian dimana sebuah kendaraan bertabrakan dengan benda lain dan menyebabkan kerusakan. 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.

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 lantas 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). 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. *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".

TINJAUAN PUSTAKA

1. Analisis Regresi

Analisis deskriptif yaitu bagian statistika mengenai pengumpulan data, penyajian, penentuan nilai-nilai statistika, pembuatan diagram atau gambar mengenai sesuatu hal (Nasution, 2017).

2. Algoritma C4.5

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). Penggunaan algoritma ini secara rekursif akan mengunjungi setiap simpul keputusan dengan tujuan untuk memilih pembagian yang optimal hingga tidak dapat dibagi lagi. Konsep yang digunakan untuk memilih nilai yang optimal, yaitu berdasarkan nilai *gain* tertinggi. Tahapan algoritma C4.5 dalam membangun pohon keputusan adalah sebagai berikut (Santosa, 2007):

1. Pemilihan atribut pembagi (*root*):

- a. Hitung *entropy* untuk semua data terhadap komposisi kelas:

$$Entropy(S) = -\sum_{i=1}^n p_i \log_2 p_i \quad (2.1)$$

dengan keterangannya sebagai berikut:

S = Himpunan kasus

n = Jumlah partisi atribut A

p_i = Jumlah kasus pada partisi ke- i

- b. Hitung *gain* untuk setiap atribut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \left| \frac{s_i}{s} \right| * Entropy(S_i) \quad (2.2)$$

dengan keterangan sebagai berikut:

S = Himpunan kasus

n = Jumlah partisi atribut A

$|s_i|$ = Jumlah kasus pada partisi ke- i

$|s|$ = Jumlah kasus dalam S

- c. Setelah nilai *gain* diperoleh, atribut dengan nilai *gain* tertinggi dipilih menjadi atribut pembagi (*root*).

2. Ulangi seluruh proses untuk setiap cabang hingga seluruh kasus untuk cabang tersebut memiliki kelas yang sama.

3. Proses partisi (pembagian) pada pohon keputusan akan berhenti setelah memenuhi salah satu kondisi berikut:

- a. Seluruh data di partisi mempunyai kelas yang sama.
- b. Tidak ada lagi atribut yang tersisa untuk membagi data.
- c. Partisi tidak memiliki data lagi.

2. Naive Bayes adalah salah satu teknik prediksi berbasis probabilistik sederhana yang didasarkan pada penerapan aturan Bayes dimana diasumsikan independensi atau ketidaktergantungan yang kuat atau naif. Dengan kata lain, dalam Naive

Bayes model yang digunakan adalah “model fitur independen”. Dalam Bayes (khususnya *Naive Bayes*), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Prasetyo, 2012). Prediksi Bayes didasarkan pada teorema Bayes dengan rumus umum sebagai berikut :

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)}$$

Penjelasan dari formula di atas sebagai berikut:

$P(H|E)$: Probabilitas akhir bersyarat (*conditional probability*) ketika suatu hipotesis H terjadi apabila diberikan bukti (*evidence*) E terjadi.

$P(E|H)$: Probabilitas ketika sebuah bukti E terjadi akan memengaruhi hipotesis H

$P(E)$: Probabilitas awal (*priori*) ketika bukti E terjadi tanpa memandang hipotesis/bukti yang lain.

$P(H)$: Probabilitas awal (*priori*) ketika hipotesis H terjadi tanpa memandang bukti apapun.

2. Adaptive Boosting (Adaboost)

Algoritma *AdaBoost* pertama kali diperkenalkan Freund dan Schapire (1995) yang telah banyak memecahkan berbagai masalah praktis dari algoritma *boosting* sebelumnya (Freund & Schapire, 1999). *Boosting* merupakan salah satu contoh metode *ensemble* (*ensemble methods*) yang menggabungkan suatu urutan model pembelajaran k (atau disebut juga sebagai *classifier* dasar), $M_1, M_2, \dots, M_k$, dengan tujuan menciptakan model klasifikasi gabungan yang lebih baik, M^* . Metode *ensemble* ini mengembalikan hasil prediksi kelas berdasarkan penilaian dari *classifier* dasarnya (Han dkk., 2012). Adapun algoritma *AdaBoost* memiliki pondasi teori yang solid, prediksi yang sangat akurat, tingkat kesederhanaan yang tinggi (cukup hanya dengan 10 baris kode) dan penggunaannya yang luas dan sukses (Wu dkk., 2007).

Adaboost dan variannya telah sukses diterapkan pada beberapa bidang (*domain*) karena dasar teorinya yang kuat, prediksi yang akurat dan kesederhanaan yang besar. Langkah-langkah pada algoritma *adaboost* adalah sebagai berikut.

1. Lakukan *input* suatu kumpulan *sample* penelitian dengan label

$\{(x_i, y_1, \dots, x_m, y_m)\}$ dengan label $y_i \in Y = \{1, \dots, k\}$, algoritma pembelajaran dasar (*base learning algorithm*) $\mathcal{L}$ dan jumlah iterasi atau

perulangan T .

2. Melakukan inisialisasi nilai bobot suatu sampel pelatihan, dengan persamaan berikut:

$$D_1(i) = \frac{1}{m} \text{ untuk } i = 1, \dots, m \quad (2.3)$$

3. Untuk $t = 1, \dots, T$
4. Kemudian latih *base learn* h_t dari sampel pelatihan menggunakan distribusi D_t
5. Hitung kesalahan dari $h_t: \epsilon_t = Pr_{i \sim D_i}[h_t(x_i) \neq y_i]$, menggunakan persamaan berikut:
 $\epsilon_t = \sum_{i: h_t(x_i) \neq y_i} D_t(i)$
 Jika $\epsilon_t \geq \frac{1}{2}$, maka set $T = t - 1$, batalkan perulangan dan langsung menuju *output*.

6. Tetapkan bobot dari h_t , dengan persamaan berikut:
 $\alpha_t \geq \frac{1}{2} \ln\left(\frac{1 - \epsilon_t}{\epsilon_t}\right)$

7. Lakukan *update* bobot sampel pelatihan, dengan persamaan berikut:

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{jika } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{jika } h_t(x_i) \neq y_i \end{cases}$$

 Dimana Z_t sebuah faktor normalisasi yang mengaktifkan $D_{t+1}(i)$ menjadi distribusi.

8. *Ouput*, dengan persamaan berikut:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(x)\right)$$

K-Fold Cross Validation

K-fold cross validation adalah salah satu metode untuk mengevaluasi kinerja *classifier*, metode ini dapat digunakan apabila memiliki jumlah data yang terbatas (jumlah *instance* tidak banyak) (Anggraeni, “Tanpa Tahun”). *K-fold cross validation* merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut *input* yang acak. *K-fold cross validation* diawali dengan membagi data sejumlah *n-fold* yang diinginkan.

Confusion Matrix

Menurut Kotu & Deshpande (2015), *Confusion Matrix* merupakan metode evaluasi yang digunakan untuk mengukur performansi pemodelan berdasarkan tabel matrix (2x2)

dengan dua nilai kelas yaitu nilai Y atau nilai N . Pada Gambar 2.3 dijelaskan bahwa nilai sel vertikal (kolom) berisi data hasil observasi sementara nilai sel horizontal (baris) berisi data prediksi

METODE PENELITIAN

1. Sumber Data

Sumber data yang digunakan dalam penelitian ini merupakan data sekunder dari Kepolisian Resor Kabupaten Pati yaitu data kecelakaan lalu lintas di Kabupaten Pati pada tahun 2019 sebanyak 207 data yang diperoleh dari Tugas Akhir Intan Kurnia Dewi Dewi (2019).

2. Variabel Penelitian

Variabel dependen bernilai 0 untuk kategori Meninggal dan 1 untuk Kategori Tidak Meninggal

Variabel independen

- X_1 : Usia Korban
- X_2 : Jenis Kelamin
- X_3 : Pendidikan
- X_4 : Jenis Pekerjaan
- X_5 : Jenis Kecelakaan
- X_6 : Jam Kejadian
- X_7 : Tanggal Kejadian
- X_8 : Alat Keselamatan
- X_9 : Peran Korban
- X_{10} : Jenis Kendaraan
- X_{11} : Faktor Manusia

3. Analisis Data

Tahapan analisis yang digunakan dalam penelitian ini adalah :

Langkah-langkah yang dilakukan pada *Adaboost C4.5* sebagai berikut:

1. Mengumpulkan data
2. Melakukan analisis deskriptif
3. Melakukan *data cleaning* yaitu membersihkan nilai-nilai dalam atribut yang kosong.
4. Membagi data menjadi data latih dan data uji
5. Memilih atribut sebagai akar
6. Membuat cabang untuk tiap-tiap nilai
7. Membagi kasus dalam cabang
8. Mengulangi proses untuk setiap cabang sampai semua kasus dalam cabang memiliki kelas yang sama
9. Melakukan *boosting* oleh *Adaboost*
10. Melakukan perulangan (iterasi) dengan *Adaboost* dibatasi sebanyak 10 kali dengan tool *Rapidminer*
11. Melakukan validasi dengan *10-fold cross validation*
 Melakukan pengujian dari tingkat Akurasi dan Spesifitas menggunakan *Confusion Matrix*

Langkah-langkah yang dilakukan pada *Adaboost Naive Bayes* sebagai berikut:

1. Mengumpulkan data
2. Melakukan Analisis deskriptif
3. Melakukan *cleaning data* yaitu membersihkan nilai-nilai dalam atribut yang kosong.
4. Membagi data menjadi data latih dan data uji
5. Menghitung peluang masing-masing atribut
6. Menghitung peluang masing-masing label
7. Menghitung peluang label bersyarat kategori atribut
8. Mengkalikan langkah 6 dengan langkah ke 4
9. Melakukan *boosting* oleh *Adaboost*
10. Melakukan perulangan (iterasi) dengan *Adaboost* dibatasi sebanyak 10 kali dengan tool *Rapidminer*
11. Melakukan validasi dengan 10-fold *cross validation*
12. Melakukan pengujian tingkat Akurasi dan Spesifitas menggunakan *Confusion Matrix*

Langkah-langkah Penelitian *Adaboost* C4.5 dan *Adaboost Naive Bayes*

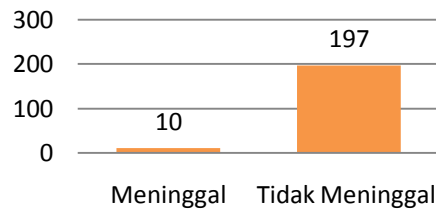
1. Mengumpulkan data
2. Melakukan Analisis Deskriptif
3. Melakukan *cleaning data* yaitu membersihkan nilai-nilai dalam atribut yang kosong.
4. Membagi data menjadi data latih dan data uji
5. Membandingkan *Adaboost* C4.5 dan *Adaboost Naive Bayes*
6. Melakukan perulangan (iterasi) dengan *Adaboost* dibatasi sebanyak 10 kali dengan tool *Rapidminer*
7. Melakukan validasi dengan 10-fold *cross validation*
8. Melakukan pengujian untuk mengukur hasil dari model dengan melihat tingkat Akurasi dan Spesifitas menggunakan *Confusion Matrix*

HASIL DAN PEMBAHASAN

1. Analisis Deskriptif

Analisis deskriptif digunakan untuk memperoleh gambaran data secara umum. Berikut adalah gambaran umum dari data sekunder kecelakaan lalu lintas di Kabupaten Pati tahun 2019.

Tingkat Keparahahan Luka Korban



Pada gambar diatas diketahui jumlah keseluruhan kasus kecelakaan lalu lintas di Kabupaten Pati tahun 2019 sebanyak 207 dengan tingkat keparahan korban kecelakaan dengan kategori meninggal sebanyak 10 orang dan kategori tidak meninggal sebanyak 197 orang. Hal tersebut menunjukkan bahwa kasus kecelakaan di Kabupaten Pati pada Tahun 2019 dengan kategori tidak meninggal lebih banyak dibandingkan dengan kasus kecelakaan dengan kategori meninggal

2. *Adaboost* C4.5

Adaptive Boosting (Adaboost) C4.5 dengan Data Latih:Data Uji 60:40

Penerapan algoritma *Adaboost* C4.5 yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel *confusion matrix* untuk data latih : data uji dengan persentase 60:40

Tingkat Keparahahan	Meninggal	Tidak Meninggal	Class Precision
Meninggal	0	0	0 %
Tidak Meninggal	10	197	95,17 %
Class Recall	0 %	100 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+197}{0+197+0+10} = \frac{197}{207} = 95,17\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{197}{197+0} = \frac{197}{197} = 100\%$$

Berdasarkan tabel diatas perhitungan akurasi dengan data latih 124 data dan data uji 83 data, 0 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 0 diprediksi sesuai tidak meninggal. Sebanyak 10 diprediksi tidak meninggal tetapi ternyata meninggal dan sebanyak 197 diprediksi sesuai tidak meninggal. Dari $TP < TN < FP < FN$ diatas juga dilakukan perhitungan *confusion matrix*, diperoleh tingkat akurasi sebesar 95,17% dan nilai Spesifitas sebesar 100%.

Adaptive Boosting (Adaboost) C4.5 dengan

Data Latih: Data Uji 70:30

Penerapan algoritma *Adaboost* C4.5 yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel *confusion matrix* untuk data latih : data uji dengan persentase 70:30

Tingkat Keparahan	Meninggal	Tidak Meninggal	Class Precision
Meninggal	0	0	0 %
Tidak Meninggal	10	197	95,17 %
Class Recall	0 %	100 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+197}{0+197+0+10} = \frac{197}{207} = 95,17\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{197}{197+0} = \frac{197}{197} = 100\%$$

Berdasarkan tabel diatas, perhitungan akurasi dengan data latih 145 data dan data uji 62 data, 0 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 0 diprediksi sesuai tidak meninggal. Sebanyak 10 diprediksi tidak meninggal tetapi ternyata meninggal dan sebanyak 197 diprediksi sesuai tidak meninggal. Dari $TP < TN < FP < FN$ diatas juga dilakukan perhitungan *confusion matrix*, diperoleh tingkat akurasi sebesar 95,17% dan nilai Spesifitas sebesar 100%.

Adaptive Boosting (Adaboost) C4.5 dengan

Data Latih: Data Uji 80:20

Penerapan algoritma *Adaboost* C4.5 yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel *confusion matrix* untuk data latih : data uji dengan persentase 80:20

Tingkat Keparahan	Meninggal	Tidak Meninggal	Class Precision
Meninggal	0	0	0 %
Tidak Meninggal	10	197	95,17 %
Class recall	0 %	100 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+197}{0+197+0+10} = \frac{197}{207} = 95,17\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{197}{197+0} = \frac{197}{197} = 100\%$$

Berdasarkan tabel diatas perhitungan akurasi dengan data latih 166 data dan data uji 41 data, 0 diklasifikasi prediksi meninggal dan

ternyata meninggal. Sebanyak 0 diprediksi sesuai tidak meninggal. Sebanyak 10 diprediksi tidak meninggal tetapi ternyata meninggal dan 197 diprediksi sesuai tidak meninggal. Dari $TP < TN < FP < FN$ diatas juga dilakukan perhitungan *confusion matrix*, diperoleh tingkat akurasi sebesar 95,17% dan nilai Spesifitas sebesar 100%.

Adaptive Boosting (Adaboost) C4.5 dengan Data Latih: Data Uji 90:10

Penerapan algoritma *Adaboost* C4.5 yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel *confusion matrix* untuk data latih : data uji dengan persentase 90:10

Tingkat Keparahan	Meninggal	Tidak Meninggal	Class Precision
Meninggal	0	0	0 %
Tidak Meninggal	10	197	95,17 %
Class recall	0 %	100 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{0+197}{0+197+0+10} = \frac{197}{207} = 95,17\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{197}{197+0} = \frac{197}{197} = 100\%$$

Berdasarkan tabel perhitungan akurasi dengan data latih 186 data dan data uji 21 data, 0 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 0 diprediksi sesuai tidak meninggal. Sebanyak 10 diprediksi tidak meninggal tetapi ternyata meninggal dan 197 diprediksi sesuai tidak meninggal. Dari $TP < TN < FP < FN$ diatas juga dilakukan perhitungan *confusion matrix*, diperoleh tingkat akurasi sebesar 95,17% dan nilai Spesifitas sebesar 100%.

Dari hasil penelitian diatas maka dapat disimpulkan pada beberapa percobaan dengan menggunakan data latih: data uji yaitu 60:40, 70:30, 80:20, 90:10 menggunakan *Adaptive Boosting (Adaboost) C4.5* pada kecelakaan lalu lintas di Kabupaten Pati pada Tahun 2019 memperoleh tingkat akurasi yang sama yaitu 95,17%.

Adaptive Boosting (Adaboost) Naive Bayes dengan Data Latih: Data Uji 60:40

Penerapan algoritma *Adaboost Naive Bayes* yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel *confusion matrix* untuk data latih : data uji dengan persentase 60:40

Tingkat Keparahan	Meninggal	Tidak Meninggal	Class Precision
----------------------	-----------	--------------------	--------------------

n		l	n
Meninggal	3	5	37,5 %
Tidak Meninggal	7	192	96,48 %
<i>Class Recall</i>	30 %	97,46 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+192}{3+197+5+7} = \frac{195}{207} = 94,2\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{197}{197+5} = \frac{197}{202} = 97,5\%$$

Berdasarkan tabel perhitungan akurasi dengan data latih 124 data dan data uji 83 data, 3 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 5 diprediksi sesuai tidak meninggal. Sebanyak 7 diprediksi tidak meninggal tetapi ternyata meninggal dan 192 diprediksi sesuai tidak meninggal. Dari TP<TN<FP<FN diatas juga dilakukan perhitungan confusion matrix, diperoleh tingkat akurasi sebesar 94,2% dan nilai Spesifitas sebesar 97,5%.

Adaptive Boosting (Adaboost) Naive Bayes dengan Data Latih:Data Uji 70:30

Penerapan algoritma *Adaboost Naive Bayes* yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel 4.6 *confusion matrix* untuk data latih : data uji dengan persentase 70:30

Tingkat Keparahan	Meninggal	Tidak Meninggal	<i>Class Precision</i>
Meninggal	3	1	75 %
Tidak Meninggal	7	196	96,55 %
<i>Class Recall</i>	30 %	99,49 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+196}{3+197+5+7} = \frac{199}{207} = 96,13\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{196}{196+1} = \frac{196}{197} = 99,49\%$$

Berdasarkan tabel 4.6, perhitungan akurasi dengan data latih 145 data dan data uji 62 data, 3 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 1 diprediksi sesuai tidak meninggal. Sebanyak 7 diprediksi tidak meninggal tetapi ternyata meninggal dan 196 diprediksi sesuai tidak meninggal. Dari TP<TN<FP<FN diatas juga dilakukan perhitungan *confusion matrix*, diperoleh tingkat

akurasi sebesar 96,13 % dan nilai Spesifitas sebesar 99,49%.

Adaptive Boosting (Adaboost) Naive Bayes dengan Data Latih:Data Uji 80:20

Penerapan algoritma *Adaboost Naive Bayes* yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel 4.7 *confusion matrix* untuk data latih : data uji dengan persentase 80:20

Tingkat Keparahan	Meninggal	Tidak Meninggal	<i>Class Precision</i>
Meninggal	3	1	75 %
Tidak Meninggal	7	196	96,55 %
<i>Class Recall</i>	30 %	99,49 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+196}{3+197+5+7} = \frac{199}{207} = 96,13\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{196}{196+1} = \frac{196}{197} = 99,49\%$$

Berdasarkan tabel 4.7, perhitungan akurasi dengan data latih 166 data dan data uji 41 data, 3 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 1 diprediksi sesuai tidak meninggal. Sebanyak 7 diprediksi tidak meninggal tetapi ternyata meninggal dan 196 diprediksi sesuai tidak meninggal. Dari TP<TN<FP<FN diatas juga dilakukan perhitungan confusion matrix, diperoleh tingkat akurasi sebesar 96,13 % dan nilai Spesifitas sebesar 99,49%.

Adaptive Boosting (Adaboost) Naive Bayes dengan Data Latih:Data Uji 90:10

Penerapan algoritma *Adaboost Naive Bayes* yang dilakukan dengan menggunakan tool *Rapidminer* menghasilkan nilai *confusion matrix*. Berikut ini adalah tabel 4.8 *confusion matrix* untuk data latih : data uji dengan persentase 90:10

Tabel 4. 1 *Adaboost Naive Bayes* dengan data latih:data uji 90:10

Tingkat Keparahan	Meninggal	Tidak Meninggal	<i>Class Precision</i>
Meninggal	3	0	30 %
Tidak Meninggal	7	197	96,57 %
<i>Class Recall</i>	30 %	100 %	

Dengan:

$$\text{Akurasi} : \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+197}{3+197+5+7} = \frac{200}{207} = 96,62\%$$

$$\text{Spesifitas} : \frac{TN}{TN+FP} = \frac{197}{197+0} = \frac{197}{197} = 100\%$$

Berdasarkan tabel 4.8, perhitungan akurasi dengan data latih 186 data dan data uji 21 data, 3 diklasifikasi prediksi meninggal dan ternyata meninggal. Sebanyak 0 diprediksi sesuai tidak meninggal. Sebanyak 7 diprediksi tidak meninggal tetapi ternyata meninggal dan 197 diprediksi sesuai tidak meninggal. Dari $TP < TN < FP < FN$ diatas juga dilakukan perhitungan *confusion matrix*, diperoleh tingkat akurasi sebesar 96,62 % dan nilai Spesifitas sebesar 100 %.

Dari hasil penelitian diatas maka dapat disimpulkan pada beberapa percobaan dengan menggunakan data latih:data uji yaitu 60:40, 70:30, 80:20, 90:10 menggunakan *Adaptive Boosting (Adaboost) Naive Bayes* pada kecelakaan lalu lintas di Kabupaten Pati pada Tahun 2019 memperoleh tingkat akurasi masing-masing 94,19%, 96,14%, 96,14%, 96,62%. Hal ini juga menunjukkan tingkat akurasi tertinggi dari percobaan yang telah dilakukan yaitu pada data latih:data uji 90:10.

Perbandingan Klasifikasi Adaboost C4.5 dan Adaboost Naive Bayes

Berikut hasil tingkat akurasi yang dihasilkan oleh tabel klasifikasi dari *tool Rapidminer* sebagai berikut:

KESIMPULAN

Simpulan

Berdasarkan hasil analisis dan pembahasan pada bab 4, maka dapat disimpulkan sebagai berikut.

1. Berdasarkan data yang diperoleh menunjukkan bahwa kasus kecelakaan Kabupaten Pati pada Tahun 2019 dengan kategori tidak meninggal lebih banyak dibandingkan dengan kasus kecelakaan dengan kategori meninggal *Adaptive Boosting (Adaboost) C4.5* pada kecelakaan lalu lintas di Kabupaten Pati pada Tahun 2019 dengan menggunakan data latih : data uji yaitu 60:40, 70:30, 80:20, 90:10 memperoleh tingkat akurasi yang sama yaitu 95,17%.
2. *Adaptive Boosting (Adaboost) Naive Bayes* pada kecelakaan lalu lintas di Kabupaten Pati pada Tahun 2019 dengan menggunakan data latih:data uji yaitu 60:40, 70:30, 80:20, 90:10 memperoleh tingkat akurasi masing-

Tabel 4. 2 *Adaboost C4.5 dan Adaboost Naive Bayes*

Data Latih : Data Uji	<i>Adaboost C4.5</i>	<i>Adaboost Naive Bayes</i>
60:40	95,17%	94,19%
70:30	95,17%	96,13%
80:20	95,17%	96,13%
90:10	95,17%	96,62%

Dari tabel 4.9 hasil perbandingan nilai kedua kerja algoritma tingkat akurasi *Adaboost Naive Bayes* dengan data latih:data uji 60:40 mempunyai tingkat akurasi 94,19 % lebih rendah dari algoritma *Adaboost C4.5* yang mempunyai tingkat akurasi 95,17%. Untuk data latih:data uji 70:30, *Adaboost Naive Bayes* mempunyai tingkat akurasi 96,13% lebih unggul dari algoritma *Adaboost C4.5* yang mempunyai tingkat akurasi 95,17%. Untuk data latih:data uji 80:20, *Adaboost Naive Bayes* mempunyai tingkat akurasi 96,13% lebih unggul dari algoritma *Adaboost C4.5* yang mempunyai tingkat akurasi 95,17%. Untuk data latih:data uji 90:10, *Adaboost Naive Bayes* mempunyai tingkat akurasi 96,62% memiliki mempunyai tingkat akurasi lebih unggul dari *Adaboost C4.5*. Hal ini menunjukkan bahwa *Adaboost Naive Bayes* pada kasus Kecelakaan Lalu Lintas di Kabupaten Pati dengan data latih:data uji 70:30, 80:20, dan 90:10 lebih baik dibandingkan dengan *Adaboost C4.5*

masing 94,19%, 96,14%, 96,14%, 96,62%.

3. Tingkat akurasi *Adaboost Naive Bayes* pada kasus Kecelakaan Lalu Lintas di Kabupaten Pati dengan data latih : data uji 70:30, 80:20, dan 90:10 lebih baik dibandingkan dengan tingkat akurasi *Adaboost C4.5*

DAFTAR PUSTAKA

- Anggraeni, D. "Tanpa Tahun", "http://lontar.ui.ac.id/," [Online]. Available: http://lontar.ui.ac.id/file?file=digital/123561-SK-739Klasifikasi%20topikAnalisis.pdf. 09 September 2020 (17:21).
- Devita, R. N., H.W. Herwanto., & A.P. Wibawa. (2018). Perbandingan Kinerja Metode Naive Bayes dan K-Nearest Neighbor untuk Klasifikasi Artikel Berbahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 5(4), 427–434. https://doi.org/10.25126/jtiik.201854773.
- Dewi, I. K. 2019. *Penerapan Algoritma Naive Bayes dalam Klasifikasi Tingkat*

- Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Pati Jawa Tengah. Akademi Statistika Muhammadiyah Semarang. Semarang.
- Fauzi, F., M.Y. Darsyah, dan T.W. Utami. 2017. Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota Se-Indonesia dengan Pendekatan *Smooth Support Vector Machine* (SSVM) Kernel Radial Basis Function (RBF). *Seminar Nasional Pendidikan, Sains dan Teknologi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Muhammadiyah Semarang*. Universitas Muhammadiyah Semarang: 88-97.
- Freund, Y. dan R.E. Schapire. 1999. A Short Introduction to Boosting. *Journal of Japanese Society for Artificial Intelligence*, 14(5), 771–780.
- Han, J., K. Micheline, dan P. Jian. 2012. *Data mining: Concepts and Techniques* (3th ed.). Waltham, MA: Elsevier/Morgan Kaufmann.
- Irawan, A. I. 2016. Penerapan Metode Naïve Bayes Classifier dan Algoritma Adaboost untuk Prediksi Penyakit Ginjal Kronik. *Skripsi*. Universitas Sebelas Maret. Surakarta.
- Jayanti, N., S. Puspitodjati., dan T. Elida, 2008. Teknik Klasifikasi Pohon Keputusan Untuk Memprediksi Kebangkrutan Bank Berdasarkan Rasio Keuangan Bank. *Proceeding Seminar Ilmiah Nasional Komputer Dan Sistem Intelijen (KOMMIT 2008) Auditorium*, (Kommit), 20–21.
- Johnson. R. A. dan Wichern. D. W. 2007. *Applied Multivariate Statistical Analysis*. Sixth Edition. New Jersey: Prentice Hall International. Inc
- Jusia, P. A. 2018. *Analisis Komparasi Pemodelan Algoritma Decision Tree Menggunakan Metode Particle Swarm Optimization dan Metode Adaboost untuk Prediksi Awal Penyakit Jantung*. 1048–1056.
- Kotu, V., B. Deshpande. 2015. *Predictive Analytic Using Rapid Miner*. Elsevier Inc. ISBN: 978-012-801460-8
- Larose, D. T. 2005. *Discovering Knowledge in Data*. New Jersey: John Willey & Sons, Inc. ISBN 0-471-66657-2.
- Mantas, C. J. dan J. Abellan. 2014. Credal-C4.5: Decision Tree Based on Imprecise Probabilities to Classify Noisy Data. *Expert System With Application*, 4625–4637.
- Nasution, L. M. 2017. STATISTIK DESKRIPTIF. *Jurnal Hikmah*, 14(1), 49–55.
- Nurzahputra, A. dan M.A. Muslim. 2017. Peningkatan Akurasi pada Algoritma C4.5 menggunakan Adaboost untuk Meminimalkan Resiko Kredit, (1), 243–247.
- Prasetyo, E. 2012. *Data Mining Konsep dan Aplikasi Menggunakan MATLAB*. Yogyakarta: ANDI Yogyakarta.
- Putranto, R. A., T. Wuryandari., dan Sudarno. 2015. Perbandingan Analisis Klasifikasi antara Decision Tree dan Support Vector Machine Multiclass untuk Penentuan Jurusan pada Siswa SMA. *Jurnal Gaussian*, 4, 1007–1016.
- Santosa, B. 2007. *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta : Graha Ilmu.
- Suprianto, W. 2019. Catatan Laka Lantas Kabupaten Pati di Akhir Tahun 2019, <http://lintaspantura.id/catatan-laka-lantas-kabupaten-pati-di-akhir-tahun-2019/>. 20 Juni 2020 (14:12).
- Syaputra, I. 2018. Klasifikasi Kejadian Kecelakaan Lalu Lintas berdasarkan Luka Korban dengan Metode Support Vector Machine Polynomial Smote (Studi Kasus: Kecelakaan Lalu Lintas di Sleman, Yogyakarta Tahun 2016-2017). *Skripsi*. Universitas Islam Indonesia (UII). Yogyakarta.
- Syfa. 2017. “Tim PPGD Satlantas Pati Siap Tolong Korban Kecelakaan Lalu Lintas”, <https://mitrapost.com/2017/04/08/tim-ppgd-satlantas-pati-siap-tolong-korban-kecelakaan-lalu-lintas/>, 23 Juni 2020 (10:00).
- Wisdayani, D. S., I.M. Nur, dan R. Wasono. 2019. Penerapan Algoritma K-Nearest Neighbor dalam Klasifikasi Tingkat Keparahan Korban Kecelakaan Lalu Lintas di Kabupaten Jawa Tengah. *Prosiding Mahasiswa Seminar Nasional*

- Unimus. *Universitas Muhammadiyah Semarang*: 373–380.
- World Health Organization (WHO). 2013. *Global Status Report On Road Safety* 2013.
- Wu, X., K. Vipin., J.R. Quinlan., J. Ghosh., Q. Yang., H. Motoda, dan G.J McLachlan. 2007. *Top 10 Algorithms in Data Mining*. *Knowledge and Information Systems*. Vol. 14. doi:10.1007/s10115-007-0114-2.
- Yoo, I., P. Alafaireet., M. Marinov., K. Hernandez., R. Gopidi., J.F. Chang dan L. Huang. 2012. *Data Mining in Healthcare and Biomedicine : A Survey of the Literature*. 2431–2448.

