

BAB II

TINJAUAN PUSTAKA

2.1 Kecelakaan

2.1.1 Pengertian Kecelakaan

Kecelakaan lalu lintas menurut UU RI Pasal 1 No. 22 tahun 2009 pasal 1 merupakan suatu peristiwa atau kejadian di jalan raya tidak diduga dan tidak disengaja yang melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan/atau kerugian harta benda. Terjadinya suatu kejadian kecelakaan selalu mengandung unsur ketidaksengajaan dan tidak disangka-sangka serta akan menimbulkan perasaan terkejut, heran dan trauma bagi orang yang mengalami kecelakaan tersebut. Apabila kecelakaan terjadi dengan disengaja dan telah direncanakan sebelumnya, maka hal ini bukan merupakan kecelakaan lalu lintas, namun digolongkan sebagai suatu tindakan kriminal baik penganiayaan atau pembunuhan yang berencana.

2.1.2 Jenis dan Bentuk Kecelakaan

Jenis dan bentuk kecelakaan dapat diklasifikasikan menjadi lima, yaitu: kecelakaan berdasarkan korban kecelakaan, kecelakaan berdasarkan lokasi kejadian, kecelakaan berdasarkan waktu terjadinya kecelakaan, kecelakaan berdasarkan posisi kecelakaan dan kecelakaan berdasarkan jumlah kendaraan yang terlibat. Penjelasan mengenai klasifikasi jenis dan bentuk kecelakaan tersebut diuraikan lebih lanjut di bawah ini Wedasana dalam Fajar (2015).

2.1.2.1 Kecelakaan Berdasarkan Korban Kecelakaan

Kecelakaan menurut korban kecelakaan menitik beratkan pada manusia itu sendiri, kecelakaan ini dapat berupa luka ringan, luka berat maupun meninggal dunia. Menurut Pasal 93 dari Peraturan Pemerintah No. 43 Tahun 1993 tentang Prasarana dan Lalu Lintas Jalan, sebagai peraturan pelaksanaan dari Undang-undang Lalu Lintas dan Angkutan Jalan, mengklasifikasikan korban dari kecelakaan sebagai berikut:

1. Kecelakaan Luka Fatal atau Meninggal

Korban meninggal atau korban mati adalah keadaan dimana korban yang dipastikan mati sebagai akibat kecelakaan lalu lintas dalam waktu paling lama 30 hari setelah kecelakaan tersebut.

2. Kecelakaan Luka Berat

Korban luka berat adalah keadaan dimana korban yang menderita cacat tetap atau harus dirawat dalam jangka waktu lebih dari 30 hari sejak terjadinya kecelakaan. Cacat tetap adalah sesuatu anggota badan hilang atau tidak dapat digunakan sama sekali dan tidak dapat sembuh atau pulih untuk selama-lamanya.

3. Kecelakaan Luka Ringan

Korban luka ringan adalah keadaan dimana keadaan korban mengalami luka-luka yang tidak membahayakan jiwa dan/atau tidak memerlukan pertolongan atau perawatan lebih lanjut di Rumah Sakit.

Dalam penelitian ini data kecelakaan dibagi menjadi dua kategori yaitu meninggal dan tidak meninggal.

2.1.2.2 Kecelakaan Berdasarkan Lokasi Kejadian

Kecelakaan dapat terjadi dimana saja disepanjang ruas jalan, baik pada jalan lurus, tikungan jalan, tanjakan dan turunan, di dataran atau di pegunungan, di dalam kota maupun di luar kota Wedasana dalam Fajar (2015).

2.1.2.3 Kecelakaan Berdasarkan Waktu Terjadinya Kecelakaan

Kecelakaan berdasarkan waktu terjadinya kecelakaan dapat digolongkan menjadi dua yaitu:

1. Jenis Hari
 - a. Hari Kerja : Senin, Selasa, Rabu, Kamis dan Jumat.
 - b. Hari Libur : Minggu dan hari-hari libur nasional.
 - c. Akhir Minggu : Sabtu.
2. Waktu
 - a. Dini Hari : Jam 00.00 – 06.00
 - b. Pagi Hari : Jam 06.00 – 12.00
 - c. Siang Hari : Jam 12.00 – 18.00
 - d. Malam Hari : Jam 18.00 – 24.00



2.1.2.4 Kecelakaan Berdasarkan Posisi Kecelakaan

Kecelakaan dapat terjadi dalam berbagai posisi tabrakan diantaranya yaitu:

1. Tabrakan pada saat menyalip (*Side Swipe*).
2. Tabrakan depan dengan samping (*Right Angle*).
3. Tabrakan muka dengan belakang (*Rear End*).
4. Tabrakan muka dengan muka (*Head On*).
5. Tabrakan dengan pejalan kaki (*Pedestrian*).

6. Tabrak lari (*Hit and Run*).
7. Tabrakan diluar kendali (*Out Of Control*).

2.1.2.5 Kecelakaan Berdasarkan Jumlah Kendaraan yang Terlibat

Kecelakaan dapat juga didasarkan atas jumlah kendaraan yang terlibat, baik kecelakaan tunggal yang dilakukan oleh satu kendaraan, kecelakaan ganda yang dilakukan oleh dua kendaraan maupun kecelakaan beruntun yang dilakukan oleh lebih dari dua kendaraan.

2.1.3 Faktor-Faktor Penyebab Kecelakaan Lalu Lintas

Kecelakaan adalah suatu kejadian yang disebabkan oleh banyak faktor, yang pada dasarnya disebabkan oleh kurang efektifnya gabungan dari faktor-faktor utama yaitu: pemakai jalan (manusia), lingkungan, jalan dan kendaraan (Harahap, 1995). Ada tiga unsur dasar yang menentukan keamanan jalan raya, yaitu: kendaraan, pengemudi serta fisik jalan itu sendiri. Untuk mengatur ketiga unsur utama tersebut diperlukan peraturan perundang-undangan, standar-standar yang mengatur syarat keamanan jalan. Untuk lebih jelas faktor-faktor tersebut diuraikan lebih lanjut di bawah ini:

2.1.3.1 Faktor Pemakai Jalan

Pemakai jalan adalah semua orang yang menggunakan fasilitas langsung dari satu jalan. Warpani (2001) menyebutkan bahwa faktor manusia sebagai pengguna jalan dapat dikategorikan menjadi dua golongan yaitu:

- a. Pengemudi, termasuk pengemudi kendaraan tak bermotor.
- b. Pejalan kaki, termasuk para pedagang asongan, pedagang kaki lima dan lain-lain.

2.1.3.2 Faktor Pengemudi

Berdasarkan Peraturan Pemerintah pasal 1 No. 44 Tahun 1993 tentang kendaraan dan pengemudi, sebagai peraturan pelaksana dari Undang-undang Lalu Lintas dan Angkutan Jalan, pengemudi adalah seseorang yang mengemudikan kendaraan bermotor atau seorang yang sedang belajar mengemudikan kendaraan bermotor. Pengemudi kendaraan baik kendaraan bermotor maupun tidak bermotor merupakan penyebab kecelakaan yang utama sehingga sangat perlu diperhatikan.

Tingkah laku pribadi pengemudi di dalam arus lalu lintas adalah faktor yang menentukan karakteristik lalu lintas yang terjadi. Bertambahnya usia atau orang yang lebih tua akan lebih banyak mengalami kecelakaan karena reflek pengemudi menjadi lebih lambat dan kemampuan fisik tertentu akan menurun (Oglesby, 1988). Faktor fisik yang penting untuk mengendalikan kendaraan dan mengatasi masalah lalu lintas adalah:

a. Penglihatan

Dari segi penglihatan manusia panca indera mata perlu mendapat perhatian besar karena hampir semua informasi dalam mengemudikan kendaraan diterima melalui penglihatan, bahkan dikatakan bahwa indera penglihatan terlalu dibebani dalam mengemudi.

b. Pendengaran

Pendengaran diperlukan untuk mengetahui peringatan-peringatan seperti bunyi klakson, sirine, peluit polisi dan lain sebagainya. Namun sering kali peringatan tersebut disertai isyarat yang dapat dilihat dengan mata. Reaksi dalam mengemudi erat hubungannya dengan kondisi fisik manusia (*Human Phisycal*

Factor), dari penerima rangsangan setelah melihat suatu tanda (rambu) sampai pengambilan tindakan tersebut terdiri dari:

1. *Perception* atau pengamatan yaitu rangsangan pada panca indera meliputi penglihatan diteruskan oleh panca indera yang lain.
2. *Identification* yaitu penelaahan atau pengidentifikasian dan pengertian terhadap rangsangan.
3. *Emotion* atau *Judgement* yaitu proses pengambilan keputusan untuk menentukan reaksi yang sesuai (misalnya, berhenti, menyalip, menepi, atau membunyikan tanda suara).
4. *Violation* (reaksi) yaitu pengambilan tindakan yang membutuhkan koordinasi dengan kendaraan, misalnya menginjak pedal rem, banting setir dan lain sebagainya.

Total waktu yang diperlukan untuk melaksanakan pengamatan (*Perception*) sampai pada reaksi (*Violation*) sering disebut PIEV Time yang besarnya = 2,5 detik, dipakai untuk menentukan jarak berhenti yang aman untuk setiap tingkat kecepatan dan PIEV Time = 2,0 detik, untuk jarak pandang di persimpangan jalan (Pignataro, 1973).

2.1.3.3 Faktor Pejalan Kaki

Pejalan kaki sebagai salah satu unsur pengguna jalan dapat menjadi korban kecelakaan dan dapat pula menjadi penyebab kecelakaan. Pejalan kaki sangat rentan mengalami cedera serius atau kematian jika ditabrak oleh kendaraan bermotor. Pelayanan terhadap pejalan kaki perlu mendapat perhatian yang optimal, yaitu dengan cara memisahkan antara kendaraan dan pejalan kaki, baik

menurut ruang dan waktu, sehingga kendaraan dan pejalan kaki berada pada tempat yang aman. Pemisahan ini dapat dilakukan dengan menyediakan fasilitas trotoar untuk mencegah agar pejalan kaki tidak berjalan secara regular di sepanjang jalan (Warpani, 2001).

Menurut WHO (2013), setiap tahun lebih dari 270.000 pejalan kaki kehilangan nyawa mereka di jalan-jalan. Kecelakaan tersebut terjadi pada saat sekolah, bekerja, menuju ibadah dan mengunjungi teman. Secara global, pejalan kaki merupakan 22% dari semua kematian jalan dan di beberapa negara proporsi ini setinggi dua pertiga. Jutaan orang terluka dalam kecelakaan lalu lintas terkait dengan pejalan kaki, beberapa berakibat cacat permanen. Kapasitas untuk merespon keselamatan pejalan kaki merupakan komponen penting dari upaya untuk mencegah kecelakaan lalu lintas jalan.

2.1.3.4 Faktor Kendaraan

Kendaraan adalah alat yang dapat bergerak di jalan, terdiri dari kendaraan bermotor dan kendaraan tidak bermotor. Menurut pasal 1 dari Peraturan Pemerintah No. 44 Tahun 1993 tentang Kendaraan dan Pengemudi, sebagai peraturan pelaksana dari Undang-undang Lalu Lintas dan Angkutan Jalan, kendaraan bermotor adalah kendaraan yang digerakkan oleh peralatan teknik yang berada pada kendaraan itu. Kendaraan bermotor dapat dikelompokkan dalam beberapa jenis, yaitu : sepeda motor, mobil penumpang, mobil bus, mobil barang dan kendaraan khusus.

Menurut Peraturan Pemerintah No. 44 Tahun 1993 tentang Kendaraan dan Pengemudi, sebagai peraturan pelaksana dari Undang-undang Lalu Lintas dan

Angkutan Jalan, setiap kendaraan bermotor harus dilengkapi dengan peralatan pengereman yang meliputi rem utama dan rem parkir dan memiliki sistem roda yang meliputi roda-roda dan sumbu roda. Roda-roda tersebut berupa pelek-pelek dan ban-ban hidup serta sumbu atau gabungan sumbu-sumbu roda yang dapat menjamin keselamatan. Selain itu sistem roda kendaraan bermotor juga harus mempunyai suspensi seperti penyangga yang mampu menahan beban, getaran dan kejutan yang dapat menjamin keselamatan dan perlindungan terhadap penggunanya. Lampu-lampu tambahan pada kendaraan bermotor bisa mengurangi resiko kecelakaan (Pignataro, 1973). Perlengkapan lampu-lampu serta alat pemantul cahaya pada kendaraan bermotor harus meliputi lampu utama dekat secara berpasangan, lampu utama jauh secara berpasangan, lampu penunjuk arah secara berpasangan di bagian depan dan bagian belakang kendaraan, lampu rem secara berpasangan, lampu posisi depan secara berpasangan, lampu mundur, lampu penerangan tanda nomor kendaraan di bagian belakang kendaraan, lampu isyarat peringatan bahaya dan lampu tanda batas secara berpasangan. Sabuk pengaman berjumlah dua atau lebih yang dipasang untuk melengkapi tempat duduk pengemudi dan tempat duduk penumpang. Sebab-sebab kecelakaan yang disebabkan oleh faktor kendaraan antara lain:

1. Kecelakaan lalu lintas yang disebabkan oleh perlengkapan kendaraan yaitu:
 - a. Alat-alat rem tidak bekerja dengan baik.
 - b. Alat-alat kemudi tidak bekerja dengan baik.
 - c. Ban atau roda dalam kondisi buruk.

- d. Tidak ada kaca spion.
2. Kecelakaan lalu lintas yang disebabkan oleh penerangan kendaraan yaitu:
 - a. Syarat lampu penerangan tidak terpenuhi.
 - b. Menggunakan lampu yang menyilaukan.
 - c. Lampu tanda rem tidak bekerja.
3. Kecelakaan lalu lintas yang disebabkan oleh pengamanan kendaraan, misalnya: Karoseri kendaraan yang tidak memenuhi syarat keamanan.
4. Kecelakaan lalu lintas yang di sebabkan oleh mesin kendaraan, contohnya: Mesin tiba-tiba mogok di jalan.
5. Karena hal-hal lain dari kendaraan, contohnya:
 - a. Muatan kendaraan terlalu berat untuk truk dan lain-lain.
 - b. Perawatan kendaraan yang kurang baik (persneling blong, kemudi patah dan lain-lain).

2.1.3.5 Faktor Jalan

Karakteristik jalan raya berkaitan dengan kegiatan lalu lalang kendaraan karena ini memiliki hubungan langsung dengan karakteristik dari pengemudi dan kendaraan (Slinn dkk., 2005). Sifat-sifat dan kondisi jalan sangat berpengaruh sebagai penyebab terjadinya kecelakaan lalu lintas. Perbaikan kondisi jalan mempengaruhi sifat-sifat kecelakaan. Ahli jalan raya dan ahli lalu lintas merencanakan jalan dengan cara yang benar dan perawatan secukupnya dengan harapan keselamatan akan didapat dengan cara demikian. Perencanaan tersebut berdasarkan pada hasil analisa fungsi jalan, volume dan komposisi lalu lintas,

kecepatan rencana, topografi, faktor manusia, berat dan ukuran kendaraan, lingkungan sosial serta dana.

Penyimpangan dari standar perencanaan dan kriteria perencanaan jalan bagi suatu ruas jalan hanya akan mengakibatkan turunnya nilai aman ruas jalan tersebut. Bila dalam pelaksanaan terpaksa menyimpang dari ketentuan standar, maka informasi atas rawan kecelakaan harus segera dipasang sebelum suatu jalan dibuka untuk umum. Selain itu pada lokasi rawan harus diberi informasi yang jelas mengenai kondisi jalan tersebut sehingga pengemudi mengetahui kondisi sekitarnya dan lebih berhati-hati. Informasi tersebut dapat berupa *delineator* (garis pembatas jalan) yang khusus digunakan pada waktu malam hari dan dilengkapi dengan cat yang dapat memantulkan cahaya tonggak di tepi jalan, mata kucing dan marka dengan cat yang dapat memantulkan cahaya.

Jalan sebagai landasan Bergeraknya kendaraan harus direncanakan sedemikian rupa agar memenuhi syarat keamanan dan kenyamanan bagi pemakainya. Perencanaan geometrik jalan harus memperhatikan : lalu lintas yang akan lewat pada jalan tersebut, kelandaian jalan, alinyemen horizontal, persilangan dan komponen pada penampang melintang (Soesantiyo, 1985).

Faktor yang disebabkan oleh faktor jalan dapat diklasifikasikan sebagai berikut:

1. Kecelakaan lalu lintas yang disebabkan karena adanya perkerasan jalan:
 - a. Lebar perkerasan yang tidak memenuhi syarat.
 - b. Permukaan jalan yang licin dan bergelombang.
 - c. Permukaan jalan yang berlubang.
2. Kecelakaan lalu lintas yang disebabkan karena adanya alinyemen jalan:

- a. Tikungan yang terlalu tajam.
 - b. Tanjakan dan turunan yang terlalu curam.
3. Kecelakaan lalu lintas yang disebabkan karena adanya pemeliharaan jalan:
- a. Jalan rusak.
 - b. Perbaikan jalan yang menyebabkan kerikil dan debu berserakan.
4. Kecelakaan lalu lintas yang disebabkan karena adanya penerangan jalan:
- a. Tidak adanya lampu penerangan jalan pada malam hari.
 - b. Lampu penerangan jalan yang rusak dan tidak diganti.
5. Kecelakaan lalu lintas yang disebabkan karena rambu-rambu lalu lintas:
- a. Rambu ditempatkan pada tempat yang tidak sesuai.
 - b. Rambu lalu lintas yang ada kurang atau rusak.
 - c. Penempatan rambu yang membahayakan pengguna jalan.

2.1.3.6 Faktor Lingkungan

Faktor lingkungan yang tergabung dalam sebagian besar hubungan kerusakan melalui jumlah struktural yang mencakup kontribusi tanah dasar dan lapisan koefisien kekuatan bawah dalam kondisi tertentu. Dengan demikian, efek dari curah hujan dan drainase berada di bawah kondisi normal tercermin dalam variabel kekuatan tersebut. Hanya ketika kondisi yang merugikan, sifat material berubah secara signifikan oleh pengaruh musim, sehingga diperlukan untuk memperkirakan negara dengan musim yang berbeda (Robinson dan Thagean, 2004).

1. Lokasi Jalan

- a. Di dalam kota, misalnya di daerah pasar, pertokoan, perkantoran, sekolah, perumahan dan lain sebagainya.
- b. Di luar kota, misalnya di daerah datar, pedesaan, pegunungan dan sebagainya.
- c. Di tempat khusus, misalnya di depan tempat ibadah, rumah sakit, tempat wisata dan lain sebagainya.

2. Iklim atau Musim

Indonesia mengalami dua macam musim yaitu musim penghujan dan musim kemarau, hal ini menjadi perhatian bagi pengemudi agar selalu waspada dalam mengemudikan kendaraannya. Selain itu adanya pergantian waktu dari pagi, siang, sore dan malam hari memberikan intensitas cahaya yang berbeda-beda. Hal tersebut mempengaruhi keadaan jalan yang terang, gelap atau remang-remang. sehingga mempengaruhi penglihatan pengemudi sewaktu mengendarai kendaraannya.

3. Volume Lalu Lintas

Arus atau volume lalu lintas pada suatu jalan raya diukur berdasarkan jumlah kendaraan yang melewati titik tertentu selama selang waktu tertentu (Oglesby, 1988). Volume lalu lintas dinyatakan dengan “Lalu lintas Harian Ratarata Pertahun” yang disebut AADT (*Average Annual Daily Traffic*) atau LHR (Lalu lintas Harian Rata-rata) bila periode pengamatan kurang dari satu tahun.

Kapasitas jalan dapat bervariasi secara substansial tergantung pada volume kendaraan atau pengguna jalan lainnya. Ini berarti model rekayasa

lalu lintas atau tindakan tidak dapat diterapkan tanpa analisis yang cermat dari kondisi setempat, misalnya berbagai kendaraan dan moda transportasi dapat melakukan evaluasi ulang mendasar dari asumsi yang mendasari dan perhitungan parameter (Robinson dan Thagean, 2004).

2.2 Analisis Deskriptif

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

2.3 Data Mining

Menurut Larose (2006) *Data mining* adalah bidang dari beberapa keilmuan yang menggabungkan teknik dari pembelajaran mesin, pengenalan pola, statistik, database dan visualisasi yang mengatasi permasalahan pengambilan informasi dari database yang besar. Menurut Gartner Group dalam Larose (2005) *data mining* merupakan suatu proses menemukan hubungan yang berarti, pola dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2006). Menurut Larose (2005) *Data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi

Terkadang peneliti secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data misalnya petugas pengumpul suara mungkin tidak menemukan keterangan

atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan gambaran suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibuat dengan *record* lengkap menjadikan nilai dari variabel independen sebagai nilai prediksi. Setelah itu, pada peninjauan berikutnya estimasi nilai dari variabel independen dibuat berdasarkan nilai variabel prediksi.

3. Prediksi

Prediksi hampir mirip dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang. Beberapa teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Pada klasifikasi, terdapat target variabel kategori. Misalnya, penggolongan pendapatan dapat dibagi menjadi tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

5. Pengklusteran

Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan yang membentuk kelas objek-objek yang memiliki kemiripan. Kluster merupakan kumpulan *record* yang mempunyai kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan

record-record dalam kluster lain. Pengklusteran berbeda dengan klasifikasi yaitu tidak adanya variabel independen dalam pengklusteran. Pengklusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel dependen. Akan tetapi, Kluster membagi dari keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan dengan *record* dalam kelompok lain akan bernilai minimal, sedangkan apabila kemiripan *record* dalam satu kelompok akan bernilai maksimal.

6. Asosiasi

Tugas asosiasi pada *data mining* yaitu menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih dikenal sebagai analisis keranjang belanja (*market basket analysis*).

2.4 Klasifikasi

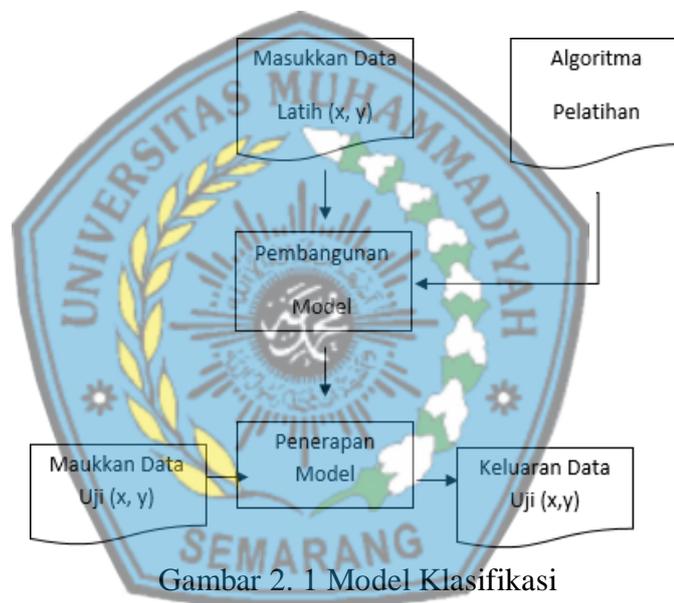
2.4.1 Konsep Klasifikasi

Klasifikasi adalah suatu pekerjaan menilai objek data untuk memasukkan ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam klasifikasi ada dua pekerjaan utama yang dilakukan, yaitu : pertama, pembangunan model sebagai *prototype* untuk disimpan sebagai memori dan kedua, penggunaan model tersebut untuk melakukan klasifikasi atau prediksi pada suatu objek data lain agar dapat mengetahui di kelas mana objek data tersebut dalam model yang mudah disimpan (Prasetyo, 2012).

2.4.2 Model Klasifikasi

Model dalam klasifikasi memiliki arti yang sama dengan kotak hitam,

dimana dalam suatu model yang menerima masukan, kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. Kerangka kerja (*framework*) klasifikasi ditunjukkan pada gambar 2.1 pada gambar tersebut disediakan sejumlah data lain (x, y) untuk digunakan sebagai data pembangunan model. Model tersebut kemudian dipakai untuk memprediksi kelas dari data uji (x, y) sehingga diketahui kelas y yang sesungguhnya.



Gambar 2. 1 Model Klasifikasi

Model yang sudah dibangun pada saat ini pelatihan kemudian dapat digunakan untuk memprediksi label kelas baru yang belum diketahui. Dalam pembangunan model saat proses pelatihan tersebut dibutuhkan suatu algoritma dalam membangunnya dikenal sebagai algoritma pelatihan. Terdapat banyak algoritma pelatihan yang sudah dikembangkan oleh para peneliti, seperti *Decision Tree*, *Naive Bayes* dan *sebagainya*. Setiap algoritma memiliki kelebihan dan kekurangan, tetapi semua algoritma memiliki prinsip sama, yaitu melakukan suatu pelatihan sehingga di akhir pelatihan, model dapat memetakan

(memprediksi) setiap vektor masukan ke label kelas keluaran dengan benar.

2.5 Klasifikasi *Decision Tree*

Decision tree merupakan metode yang sangat kuat dan populer untuk melakukan klasifikasi dan prediksi. Penggunaan berbasis *decision tree* disebabkan oleh fakta bahwa pohon keputusan dapat mewakili aturan sehingga dapat dipahami oleh manusia. Berdasarkan aturan tersebut maka dapat diekspresikan dalam bentuk *database* yang dapat digunakan untuk membagi catatan yang besar menjadi kumpulan catatan yang lebih kecil dan terurut dengan menerapkan urutan aturan keputusan sederhana (Berry dan Linoff, 2004).

Model dari *decision tree* terdiri dari seperangkat aturan untuk membagi populasi heterogen yang besar menjadi kelompok yang lebih kecil dan lebih homogen dan sehubungan dengan variabel target tertentu. Variabel *target* pada *decision tree* biasanya digunakan untuk menghitung probabilitas untuk mengklasifikasikan catatan ke dalam kelas yang paling memungkinkan.

Algoritma pembentuk *decision tree*, diantaranya adalah ID3, CART, dan C4.5. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 (Lorase, 2005).

2.5.1 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:

$$Entropy(S) = \sum_{i=0}^n -p_i * \log_2 p_i \quad (2.1)$$

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:

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

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.

2.6 Klasifikasi *Naive Bayes*

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.

2.6.1 *Naive Bayes* untuk Klasifikasi

Penelitian ini menggunakan *Naive bayes* karena dalam proses klasifikasi dalam perhitungan probabilistik, *naive bayes* lebih banyak keuntungannya. Salah satu kelebihan adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu class. *Naive Bayes* didasarkan pada teorema *Bayes* yang mempunyai kemampuan klasifikasi serupa dengan decision tree dan neural network. *Naive Bayes* terbukti mempunyai akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Selain itu berikut kelebihan yang terdapat pada *naive bayes* secara keseluruhan, yaitu:

1. Menangani kuantitatif dan data diskrit
2. Kokoh untuk titik noise yang diisolasi, dimisalkan titik yang dirata-ratakan ketika mengestimasi peluang bersyarat data.
3. Hanya membutuhkan sejumlah kecil data pelatihan untuk mengestimasi parameter (rata-rata dan variansi dari variabel) yang dibutuhkan untuk klasifikasi.
4. Menangani nilai yang hilang dengan mengabaikan instansi selama perhitungan estimasi peluang
5. Cepat dan efisiensi ruang
6. Kokoh terhadap atribut yang tidak relevan

Kaitan antara *Naive Bayes* dengan klasifikasi, korelasi hipotesis dan bukti klasifikasi adalah bahwa hipotesis dalam teorema *Bayes* merupakan label kelas

yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi. Jika X adalah vektor masukan yang berisi fitur dan Y adalah label kelas, *Naive Bayes* dituliskan dengan $P(X|Y)$. Notasi tersebut diartikan sebagai probabilitas label kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini diartikan sebagai probabilitas akhir (*posterior probability*) untuk Y , sedangkan $P(Y)$ disebut probabilitas awal (*prior probability*) Y .

Selama proses pelatihan perlu dilaksanakan pembelajaran probabilitas akhir $P(Y|X)$ pada model untuk setiap kombinasi X dan Y yang berdasarkan informasi yang diperoleh dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari Y' dengan memaksimalkan nilai $P(X'|Y')$ yang didapat.

Formulasi *Naive Bayes* untuk klasifikasi adalah :

$$P(Y, X) = \frac{P(y) \prod_{i=1}^q P(X_i' | Y)}{P(x)}$$

$P(Y|X)$ yaitu probabilitas data dengan vektor X pada kelas Y .

$P(y) \prod_{i=1}^q P(X_i' | Y)$ yaitu Probabilitas independen kelas Y dari semua fitur dalam vektor X . Nilai $P(X)$ selalu tetap sehingga dalam perhitungan prediksi nantinya kita tinggal menghitung bagian $P(y) \prod_{i=1}^q P(X_i' | Y)$ dengan memilih yang terbesar sebagai kelas yang dipilih sebagai hasil prediksi. Sementara probabilitas independen $\prod_{i=1}^q P(X_i' | Y)$ tersebut merupakan pengaruh semua fitur dari data terhadap setiap kelas Y

2.6.2 Karakteristik *Naïve Bayes*

Klasifikasi dengan menggunakan *Naïve Bayes* bekerja didasarkan pada teori probabilitas yang menganggap semua fitur dari data sebagai bukti dalam probabilitas. Hal ini memberikan karakteristik *Naïve Bayes* sebagai berikut:

1. Metode ini bekerja teguh (*robust*) terhadap data-data yang terisolasi yang biasanya merupakan data dengan karakteristik berbeda (*outliner*). N juga bisa menangani nilai atribut yang salah dengan mengabaikan data latih selama proses pembangunan model dan prediksi.
2. Tangguh menghadapi atribut yang tidak relevan
3. Atribut yang mempunyai korelasi bisa mendegradasi kinerja klasifikasi *Naïve Bayes* karena asumsi independensi atribut tersebut sudah tidak ada.

2.7 *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 $h_t: \epsilon_t = Pr_{i \sim D_t}[h_t(x_i) \neq y_i]$, 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:

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

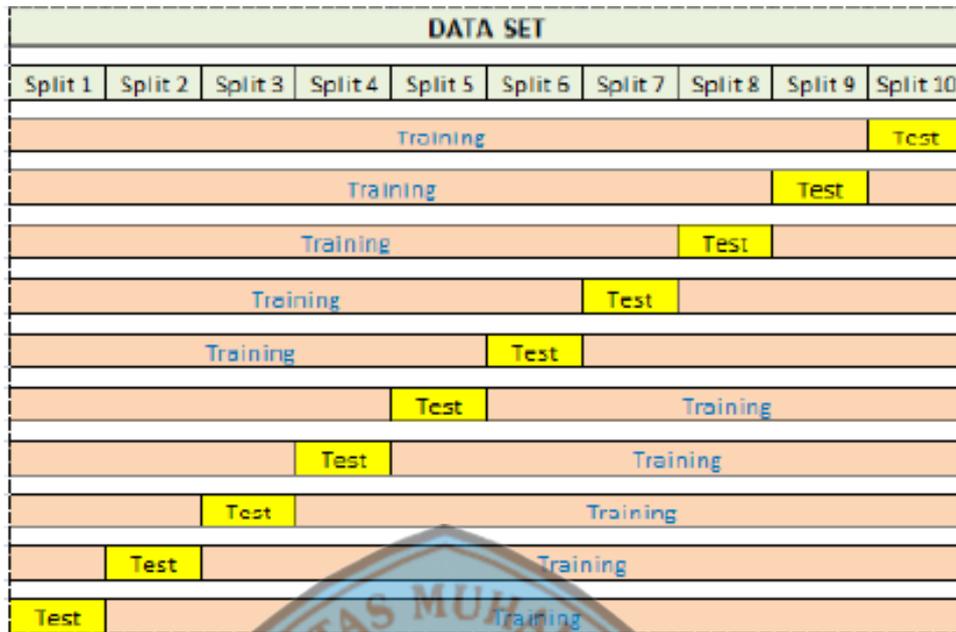
Dimana z_t sebuah faktor normalisasi yang mengaktifkan $D_{t+1}(i)$ menjadi distribusi.

8. *Ouput*, dengan persamaan berikut:

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

2.8 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. Dalam proses *cross validation data* akan dibagi dalam *n* buah partisi dengan ukuran yang sama $D_1, D_2, D_3 \dots D_n$ selanjutnya proses uji dan latih dilakukan sebanyak *n* kali. Dalam iterasi ke-*i* partisi D_i akan menjadi data uji dan sisanya akan menjadi data latih. Untuk penggunaan jumlah *fold* terbaik untuk uji validitas, dianjurkan menggunakan *10-fold cross validation* dalam model (Kohavi, 1995) Contoh pembagian dataset dalam proses *10-fold cross validation* terlihat pada Gambar 2.2



Sumber : Han dan Kamber, 2007

Gambar 2. 2 K-Fold Cross Validation

Cara kerja *K-fold cross validation* adalah sebagai berikut:

1. Total instance dibagi menjadi N bagian.
2. *Fold* ke-1 adalah ketika bagian ke-1 menjadi data uji (*testing data*) dan sisanya menjadi data latih (*training data*). Selanjutnya, hitung akurasi berdasarkan porsi *data* tersebut. Perhitungan akurasi tersebut dengan menggunakan persamaan sebagai berikut :

$$Akurasi = \frac{\sum \text{data uji benar klasifikasi}}{\sum \text{total data uji}} \times 100\%$$

3. *Fold* ke-2 adalah ketika bagian ke-2 menjadi data uji (*testing data*) dan sisanya menjadi data latih (*training data*). Selanjutnya, hitung akurasi berdasarkan porsi data tersebut.

4. Demikian seterusnya hingga mencapai *fold* ke-K. Hitung rata-rata akurasi dari K buah akurasi di atas. Rata-rata akurasi ini menjadi akurasi final.

2.9 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. Penentuan tingkat akurasi dikelompokkan menjadi 4 cara yang berbeda yaitu :

- a. Jika nilai observasi positif sesuai dengan nilai ekspektasi yang ditentukan, maka hasilnya adalah *True Positive* (TP).
- b. Jika nilai observasi negatif namun masih dikategorikan sesuai dengan nilai ekspektasi, maka hasilnya adalah *False Positive* (FP).
- c. Jika nilai observasi positif tetapi tidak sesuai dengan nilai ekspektasi, maka hasilnya adalah *False Negative* (FN).
- d. Nilai observasi yang negatif dan dinyatakan tidak sesuai dengan nilai ekspektasi, maka hasilnya adalah *True Negative* (TN).

	Observation Value		
		Y	N
Expectation Value	Y	True Positive	False Positive
	N	False Negative	True Negative

Gambar 2. 3 *Confusion Matrix*

Berdasarkan nilai True Negatif (TN), False Positive (FP), False Negatif (FN) dan True Positive (TP) dapat diperoleh tingkat akurasi, presisi dan recall.

a. Akurasi

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual.

Dengan rumus sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

b. Spesifisitas

Spesifisitas adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Dengan rumus sebagai berikut :

$$\text{Spesifisitas} = \frac{TN}{TN+FP}$$

2.10 *Rapidminer*

RapidMiner merupakan software atau perangkat lunak untuk pengolahan data, dengan menggunakan prinsip dan algoritma data mining, *Rapidminer* mengekstrak pola-pola dari data set yang besar dengan mengkombinasikan metode statistika, kecerdasan buatan dan database.

Rapidminer memudahkan penggunaanya dalam melakukan perhitungan data yang sangat banyak dengan menggunakan operator-operator. Operator ini berfungsi untuk memodifikasi data. Data dihubungkan dengan node-node pada operator kemudian pengguna hanya tinggal menghubungkannya ke node hasil untuk melihat hasilnya. Hasil yang diperlihatkan *RapidMiner* pun dapat ditampilkan secara visual dengan grafik menjadikan *RapidMiner* adalah salah satu software pilihan untuk melakukan ekstraksi data dengan metode-metode data mining (Rahmat, 2017)