

**KLASIFIKASI TINGKAT KEPADATAN LALU LINTAS OLEH
PENGGUNA KENDARAAN BERMOTOR DENGAN
MENGUNAKAN *RANDOM FOREST*
(STUDI KASUS: JALAN RAYA *BY PASS* JOMIN CIKAMPEK)**

TESIS

Disusun sebagai salah satu syarat untuk
Memperoleh gelas Magister Komputer
Dari Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI

Oleh:

HADI KUSUMAH

NPM: 2019210012



**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2020**

**KLASIFIKASI TINGKAT KEPADATAN LALU LINTAS OLEH
PENGGUNA KENDARAAN BERMOTOR DENGAN
MENGUNAKAN *RANDOM FOREST*
(STUDI KASUS: JALAN RAYA *BY PASS* JOMIN CIKAMPEK)**

Oleh:

HADI KUSUMAH

NPM: 2019210012

Bandung, 5 November 2020
Menyetujui,

Dr. Djajasukma Tjahjadi., S.E.,M.T
Pembimbing

**PROGRAM STUDI PASCASARJANA
MAGISTER SISTEM INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER LIKMI
BANDUNG
2020**

KATA PENGANTAR

Puji syukur kita panjatkan kepada dzat yang maha kekal Tuhan semesta alam yakni Allah SWT dengan segala limpahan rahmat dan hidayahnya penulis dapat menyelesaikan penulisan tesis ini dengan mengambil judul “Klasifikasi Tingkat Kepadatan Lalu Lintas Oleh Pengguna Kendaraan Bermotor Dengan Menggunakan *Random Forest*”. Penelitian ini dibuat sebagai syarat untuk mendapatkan gelar Magister Komputer di Sekolah Tinggi Manajemen dan Informatika LIKMI.

Dalam penulisan tesis ini tidak sedikit yang sangat berperan penting keterlibatannya untuk menyelesaikan penulisan tesis, maka dengan itu penulis mengucapkan terima kasih yang sebesar-besarnya kepada :

1. Bapak Dr. Djajasukma Tjahjadi., S.E., M.T yang telah dengan sabar memberikan bimbingan, masukan dan saran serta panduan dan waktu selama proses penulisan tesis.
2. Teman-teman kelas M angkatan 2019 STMIK LIKMI yang super dan saling dukung dalam segala hal baik, pikiran, waktu, tempat dan bantuan moral yang kuat agar tetap selesai bersama.
3. Rekan-rekan para peneliti bidang lalu lintas PUSJATAN yang telah dengan sabar berdiskusi dan memberi banyak pencerahan.
4. Keluarga yang selalu dicintai dan di rahmati Allah SWT.

Penulis menyadari bahwa pada penulisan tesis ini masih jauh dari sempurna karena dengan segala keterbatasan pengetahuan, waktu yang dimiliki oleh penulis. Oleh karena itu diharapkan masukan kritik dan saran dari segenap pembaca untuk mengembangkan kembali ilmu pengetahuan yang sudah dilakukan oleh penulis.

Bandung, 10 Agustus 2020

Penulis

DAFTAR ISI

KATA PENGANTAR	ii
DAFTAR ISI	iii
DAFTAR GAMBAR DAN ILUSTRASI	v
DAFTAR TABEL	vi
DAFTAR RUMUS	vii
DAFTAR LAMPIRAN	viii
ABSTRAK.....	ix
ABSTRACK.....	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Ruang Lingkup Penelitian.....	3
1.5 Metodologi Penelitian	4
1.6 Sistematika Penulisan	4
BAB II LANDASAN TEORI	6
2.1 <i>Data Mining</i>	6
2.2 <i>Machine Learning</i>	11
2.3 <i>Decision Tree</i>	11
2.4 Pengertian <i>Random Forest (RF)</i>	12
2.5 <i>RapidMiner Studio</i>	14
2.6 Lalu Lintas	15
2.6.1 <i>Volume Capacity Ration</i>	16
2.6.2 Kapasitas Jalan	16
2.6.3 Volume Lalu Lintas.....	17

2.7	Penelitian Terkait	18
BAB III OBJEK DAN METODE PENELITIAN		21
3.1	Kepadatan Lalu Lintas	21
3.2	Metode Penelitian	23
3.2.1	Pengumpulan Data	24
3.2.2	Metode Analisis	24
3.3	Persiapan Data (<i>Data Preparation</i>)	25
BAB IV HASIL PENELITIAN DAN PEMBAHASAN		27
4.1	<i>Data Cleaning</i>	27
4.3	<i>Data Transformation</i>	29
4.4	<i>Data Mining</i>	33
4.4.1	Analisis Deskriptif	33
4.4.2	<i>Random Forest</i>	36
4.4.2	<i>Performance Random Forest</i>	41
BAB V KESIMPULAN DAN SARAN		42
5.1	Kesimpulan	42
5.2	Saran	42
DAFTAR PUSTAKA		44

DAFTAR GAMBAR DAN ILUSTRASI

Gambar 2.1 Proses <i>Data Mining</i>	6
Gambar 2.2 <i>Random Forest</i>	13
Gambar 2.2 Jendela <i>Rapidminer</i>	14
Gambar 3.1 Komponen Sistem	21
Gambar 3.2 Ilustrasi Cara Kerja Sistem	22
Gambar 3.3 Ilustrasi Lalu lintas di jalan raya <i>By Pass Jomin</i>	22
Gambar 3.4 Diagram Alir Penelitian	23
Gambar 4.1 Survei Jalan Untuk Mendapatkan Nilai Kapasitas Jalan	29
Gambar 4.2 Mengamati Untuk Mendapat Nilai Hambatan Samping	30
Gambar 4.3 <i>View Data</i> Lalu Lintas Pada <i>Software Rapidminer</i>	34
Gambar 4.4 Total jumlah kendaraan per hari dari data lalu lintas yang di proses	34
Gambar 4.5 Grafik Jumlah Total kendaraan Per Hari Dari Setiap golongan	35
Gambar 4.6 Jumlah rata-rata kendaraan lewat dalam 5 menit di hari tertentu	35
Gambar 4.7 Grafik perbandingan kapasitas jalan dengan volume lalu lintas	36
Gambar 4.8 <i>Tree</i> Nilai VCR Menentukan Macet	36
Gambar 4.9 <i>Tree</i> Nilai Y Menentukan Macet	37
Gambar 4.10 <i>Tree</i> Nilai Volume Lalu Lintas	37
Gambar 4.11 <i>Tree</i> Jam 05:30	38
Gambar 4.12 <i>Tree</i> Tingkat Kepadatan Lalu lintas	38
Gambar 4.13 <i>Tree</i> Golongan I Dengan VCR	39
Gambar 4.14 <i>Tree</i> Golongan 2 dengan Volume Lalu Lintas	40
Gambar 4.15 <i>Tree</i> Golongan 2 Dengan Y	40
Gambar 4.16 <i>Tree</i> Jam Dengan Y	41
Gambar 4.17 <i>Performance Random Forest</i>	41
Gambar 4.18 Komparasi <i>Random Forest</i>	42

DAFTAR TABEL

Tabel 2.1 Kriteria Tingkat Pelayanan	16
Tabel 2.2 Kriteria Tingkat Pelayanan	18
Tabel 2.3 Perbandingan Penelitian Terkait	19
Tabel 3.1 Struktur Tabel Data Lalu Lintas	25
Tabel 4.1 Data Lalu Lintas Pada Excel	27
Tabel 4.2 Struktur Tabel Setelah Data Kecepatan Di Hapus	28
Tabel 4.3 Struktur Tabel Setelah Mencantumkan Nilai Kapasitas Jalan	30
Tabel 4.4 Struktur Tabel Setelah Mencantumkan Nilai VCR dan Kondisi Lalu Lintas	31
Tabel 4.5 Struktur Tabel Setelah Mencantumkan Nilai VCR dan Kondisi Lalu Lintas	32

DAFTAR RUMUS

Rumus 2.1 Volume Kapasitas Rasio	16
Rumus 2.2 Kapasitas Jalan	16
Rumus 2.3 Arus Lalu Lintas	17
Rumus 2.4 Volume Kendaraan Bermotor	18

DAFTAR LAMPIRAN

Lampiran A Peta Lokasi Jalan Raya <i>By Pass</i> Jomin	47
Lampiran B Contoh <i>Dataset</i>	51

ABSTRAK
Klasifikasi Tingkat Kepadatan Lalu Lintas Oleh Pengguna Kendaraan Bermotor Dengan Menggunakan *Random Forest*
(STUDI KASUS: JALAN RAYA BY PASS JOMIN CIKAMPEK)

Hadi Kusumah
Program Studi Magister Sistem Informasi
Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI
Email : hadytkom@gmail.com

Meningkatnya jumlah kepemilikan kendaraan bermotor, mengakibatkan terjadinya peningkatan pengguna jalan sehingga arus kendaraan yang lewat telah melampaui kapasitas jalan. Ketika jalan telah melampaui kapasitasnya menyebabkan kemacetan bahkan Insiden. Untuk mengatasi masalah tersebut maka diperlukan manajemen lalu lintas yang terencana dan terarah sehingga solusi pada satu titik tidak akan mengakibatkan masalah pada titik yang lain.

Untuk manajemen lalu lintas yang terencana dan terarah, terlebih dahulu perlu diketahui karakteristik lalu lintas seperti volume lalu lintas dan kapasitas jalan. Dengan mengetahui volume lalu lintas dan kapasitas maka dapat diketahui tingkat pelayanan dari ruas jalan tersebut.

Dalam mendapatkan data lalu lintas harus dilakukan survei langsung di lokasi, selain manual survei juga dapat menggunakan perangkat pencacah lalu lintas elektronik yang akan otomatis menghitung jumlah kendaraan dan mengirim data geometri jalan ke server perangkat tersebut. Penelitian ini mengklasifikasikan data yang terkumpul pada perangkat pencacah lalu lintas elektronik dengan *data mining* menggunakan *Random Forest* untuk lokasi jalan *By Pass* Jomin Cikampek.

Hasil penelitian menyimpulkan bahwa pada jalan tersebut tidak macet pada hari Selasa, Rabu, Kamis, Sabtu namun pada jam-jam aktivitas puncak nilai VCR (*Volume Capacity Rasio*) mendekati angka 1 dan pada hari Senin pagi, Jumat sore, Minggu sore terdapat sedikit kemacetan ini terjadi dikarenakan jalan tersebut merupakan jalan luar kota dan merupakan titik pertemuan jalan dari berbagai arah. Tingkat akurasi *Random Forest* pada data lalu lintas ini mempunyai akurasi sebesar 95,6%, tingkat *error* didapatkan sebesar 4,4% dan nilai *recall* 100%.

Kata Kunci: *Data mining, Lalulintas, Kemacetan, Kendaraan bermotor, Klasifikasi, Random Forest*

ABSTRACT
Classification of Traffic Congestion Level due to Vehicle Users using
Random Forest
(CASE STUDY: BY PASS JOMIN CIKAMPEK ROAD)

Hadi Kusumah
Master Program in Information Systems
Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI
Email : hadytkom@gmail.com

The increased number of vehicle ownership has resulted in increased number of road users so the flow of passing vehicles has exceeded the road capacity. When the road capacity has exceeded, there will be congestion and even incidents. To overcome this problem, a planned and directed traffic is needed so the solution used at one point or area will not cause other problem at different point or area.

For planned and directed traffic management, it is necessary to first know the traffic characteristics of the road, especially traffic volume and road capacity. By knowing the traffic volume and capacity, we will know the service level of these roads.

In obtaining traffic data, a direct survey must be carried out on the spot. Apart from manual survey, an electronic traffic counter device can also be used which will automatically count the number of vehicles and send the road geometry information to the server. This study classifies the data collected by an electronic traffic counter with data mining using Random Forest method. The location used in this study is By Pass Jomin Cikampek road.

The results of the study conclude that the road is not congested on Tuesday, Wednesday, Thursday, Saturday, but during peak activity hours the VCR (Volume Capacity Ratio) value is close to 1 and on Monday morning, Friday afternoon, Sunday afternoon there is a little traffic jam. This behaviour happened due to the road is a road outside the city and is the meeting point of the road from various directions. The Random Forest method in this traffic data study has an accuracy of 95.6%, the error rate of 4.4% and the recall value 100% .

Keyword: *Data mining, Traffic flow, Congestion, Vehicle, Classification, Random forest*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Meningkatnya jumlah kepemilikan kendaraan bermotor, mengakibatkan terjadinya peningkatan pengguna jalan, arus kendaraan yang lewat telah melampaui kapasitas jalan. Ruas jalan akan mengalami kemacetan, antrean atau tundaan serta berpeluang terjadi kecelakaan lalu lintas yang dapat mengganggu kelancaran dan kenyamanan berkendara. Karena hal ini maka menimbulkan permasalahan transportasi mulai dari jalan, alat transportasi, hingga infrastruktur penunjang transportasi. sehingga menyebabkan terjadinya penurunan tingkat pelayanan jalan yang disebabkan penggunaan ruang jalan yang tidak sebagaimana mestinya.

Populasi sepeda motor di luar Asia Tenggara sangat sedikit digunakan di jalan kota maupun antar kota dengan alasan bahwa sepeda motor tidak aman dipakai sebagai mode transportasi sehari-hari. Alasan lain mengapa teknologi penghitung volume lalu lintas dari luar Asia Tenggara kurang cocok untuk digunakan di Indonesia karena para pengembang teknologi melakukan pendekatan perhitungan kendaraannya dengan metode kendaraan bergerak mengikuti antrian per lajunya, sehingga beberapa perangkat lunak luar tidak mengenal adanya kendaraan yang bergerak secara paralel di dalam satu lajur. Sementara untuk kondisi di Indonesia adalah karakter pengendara bergerak mengikuti celah kosong di dalam arah perjalanannya. (Hanafiah. 2018)

Jalan raya *By pass* Jomin Cikampek merupakan titik yang sering terjadinya kemacetan sampai puluhan jam. Namun setelah pembangunan jalan Tol Cikopo ke Palimanan atau Cipali jalan ini tidak macet karena sebagian kendaraan memilih jalan Tol Cikopo (Detik.com. 2016). Jalan raya ini pertemuan arus kendaraan dari Purwakarta, pintu Tol Cikampek dan arah jalan arteri Karawang, berjalannya waktu karena pembangunan kawasan industri semakin luas arus kendaraan semakin ramai dan kadang terjadi kemacetan.

Untuk mengatasi masalah tersebut maka diperlukan manajemen lalu lintas yang terencana dan terarah sehingga solusi pada satu titik tidak akan mengakibatkan masalah pada titik yang lain. Untuk manajemen lalu lintas yang terencana dan terarah, terlebih dahulu perlu diketahui perilaku karakteristik lalu lintas seperti volume dan kapasitas. Dengan mengetahui volume dan kapasitas maka dapat diketahui berapa tingkat pelayanan dari ruas jalan tersebut. Setelah didapatkan informasi tersebut maka dapat mengklasifikasikan tingkat kepadatan lalu lintas pengguna kendaraan bermotor.

Alat pencacah lalu lintas otomatis dapat menghitung kendaraan dan membedakan golongan kendaraan tersebut, alat ini dapat menghitung lalu lintas kendaraan setiap waktu, *raw data* hasil *monitoring* dikirim ke server setiap 5 menit sehingga data yang ada cukup melimpah.

Berdasarkan uraian di atas diperlukan analisis untuk mengklasifikasikan data tingkat kepadatan lalu lintas yang jumlahnya banyak. Pada permasalahan ini dapat diterapkan teknik *data mining* untuk melakukan klasifikasi. Metode klasifikasi *Machine Learning* telah lama digunakan di bidang *data mining*. Menurut Han (2012), klasifikasi adalah proses menemukan kumpulan pola data satu dengan yang lainnya untuk dapat digunakan memprediksi data yang belum memiliki kelas data tertentu. Sebagai *algoritma* klasifikasi yang melibatkan metode dari statistik, kecerdasan buatan dan manajemen *database*, juga dapat di kategorikan sebagai kunci elemen dalam interpretasi data dan visualisasi data. *Decision tree* atau pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal, selain itu juga berguna untuk mengeksplorasi data dan menemukan hubungan yang tersembunyi.

Akurasi merupakan ukuran kinerja ketepatan klasifikasi *data mining*, untuk melakukan klasifikasi ini dapat menggunakan *Random forest* yang sudah dibuktikan oleh para peneliti mempunyai akurasi tinggi dalam klasifikasi karena *Random forest* merupakan metode pohon gabungan yang berasal dari pengembangan metode *Classification and Regression Tree (CART)*, yaitu dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection*. *Random forest* juga telah

diaplikasikan pada berbagai permasalahan dalam bidang kesehatan, bisnis, pendidikan dan lain-lain

Berdasarkan hasil pengamatan yang terjadi di atas, maka penulis tertarik mengambil judul penelitian "KLASIFIKASI TINGKAT KEPADATAN LALU LINTAS OLEH PENGGUNA KENDARAAN BERMOTOR MENGGUNAKAN *RANDOM FOREST*".

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah diuraikan di atas, yang menjadi permasalahan dalam penelitian ini adalah:

1. Bagaimana menerapkan klasifikasi tingkat kepadatan lalu lintas menggunakan *Random Forest* dan bagaimana hasil klasifikasinya?
2. Bagaimana mengukur keberhasilan *Random Forest* untuk pengklasifikasian data?
3. Bagaimana membuat bahan referensi tambahan untuk *stakholder* untuk manajemen lalu lintas yang lebih terencana dan terarah?

1.3 Tujuan Penelitian

Tujuan dalam melakukan penelitian ini adalah:

1. Mengklasifikasi tingkat kepadatan lalu lintas menggunakan *Random Forest*
2. Mengetahui tingkat akurasi pohon keputusan yang dibentuk *Random Forest*
3. Hasil penelitian sebagai bahan referensi tambahan *stakholder* untuk manajemen lalu lintas yang lebih terencana dan terarah

1.4 Ruang Lingkup Penelitian

Mengingat permasalahan yang dikaji sangat luas, agar penyajian lebih terarah dan mencapai sasaran yang ditentukan, maka diperlukan suatu pembatasan masalah atau ruang lingkup kajian yang meliputi hal-hal sebagai berikut:

1. Ruas jalan yang akan diteliti salah satu ruas jalan *By Pass* Jomin Cikampek
2. *Datasheet* yang dipakai diambil dari Perangkat pencacah lalu lintas elektronik hasil penelitian Pusat Penelitian Jalan dan Jembatan (PUSJATAN)
3. Klasifikasi tingkat kepadatan berdasarkan data volume dan kapasitas
4. Klasifikasi tingkat kepadatan lalu lintas menggunakan *Random Forest*

1.5 Metodologi Penelitian

Metode penelitian dalam penyusunan penelitian ini dibagi menjadi dua tahap yaitu, tahap pengumpulan data dan tahap analisa.

1. Metode Pengumpulan Data

Pengumpulan data dengan cara mencari dan mempelajari bermacam-macam *literatur* yang dibutuhkan, baik dari buku, jurnal, *paper* dan bacaan-bacaan yang ada kaitannya dengan judul penelitian. Observasi: Pengumpulan data dengan cara melakukan pengamatan langsung kondisi lalu lintas di lapangan.

2. Metode Analisis

Merupakan tahap menganalisis hal-hal yang diperlukan dalam pelaksanaan penelitian. Pada tahap ini juga dilakukan analisis terhadap pohon keputusan yang dibangun *Random Forest*.

1.6 Sistematika Penulisan

Sistematika penulisan penelitian ini disusun untuk memberikan gambaran umum tentang penelitian yang dijalankan. Sistematika penulisan tesis ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, maksud dan tujuan, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi landasan teori yang membahas tentang *Data Mining*, Lalu lintas, Klasifikasi, *Random Forest (RF)* yang digunakan untuk mengerjakan penelitian.

BAB III OBYEK DAN METODE PENELITIAN

Bab ini berisi analisis masalah, definisi kebutuhan perangkat lunak, penjelasan mengenai tahap analisis klasifikasi kepadatan lalu lintas kendaraan menggunakan *Random Forest* yang digunakan.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Dalam bab ini menjelaskan tentang hasil-hasil analisis yang telah diteliti dan pembahasan tentang hasil tersebut.

BAB V KESIMPULAN DAN SARAN

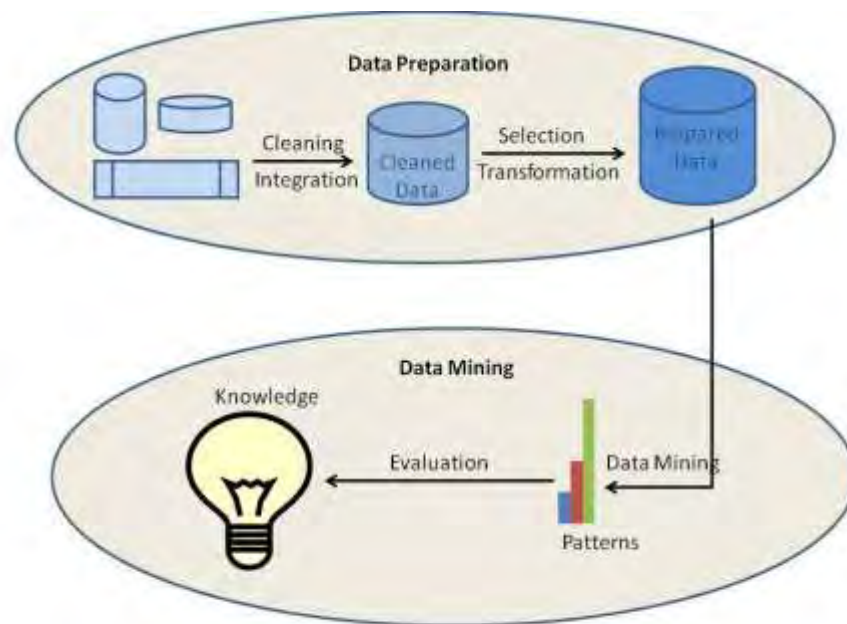
Bab ini memberikan kesimpulan dari hasil penelitian yang telah dilakukan. Serta saran-saran yang dapat membantu pembaca atau peneliti selanjutnya

BAB II

LANDASAN TEORI

2.1 Data Mining

Data mining menurut Saleh (2015), *Data mining* merupakan ekstraksi atau pemahaman *pattern* yang menarik pada data, atau dapat juga diartikan sebagai rangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual.



Sumber: www.wideskills.com

Gambar 2.1 Proses Data Mining

Pada gambar 2.1 menunjukkan proses penambangan data yang berbeda dapat diklasifikasikan menjadi dua jenis: persiapan data atau pemrosesan awal data dan penambangan data. Padahal, empat proses pertama yaitu pembersihan data, integrasi data, pemilihan data dan transformasi data dianggap sebagai proses persiapan data. Tiga proses terakhir termasuk *data mining*, evaluasi pola dan representasi pengetahuan diintegrasikan ke dalam satu proses yang disebut *data mining*. Proses *data mining* terdiri dari langkah-langkah sebagai berikut (Han, Kamber, & Pei, 2012):

1. *Data Cleaning*

Data cleaning adalah proses di mana data dibersihkan. Data di dunia nyata biasanya tidak lengkap, terdapat *noise* dan tidak konsisten. Data yang tersedia di sumber data mungkin kurang nilai atribut, data yang diminati dan lain-lain. Misalnya, butuh data pelanggan sedangkan data yang tersedia tidak menyertakan atribut jenis kelamin atau usia pelanggan, Maka datanya tentu saja tidak lengkap. Terkadang data mungkin berisi kesalahan atau menyimpang. Contohnya adalah atribut usia dengan nilai 200. Jelas bahwa nilai usia salah dalam kasus ini. Datanya juga bisa tidak konsisten. Misalnya, nama seorang karyawan mungkin disimpan secara berbeda dalam tabel data atau dokumen yang berbeda. Di sini, datanya tidak konsisten. Jika datanya tidak bersih, hasil *data mining* tidak akan andal atau akurat. *Data cleaning* melibatkan sejumlah teknik termasuk pengisian nilai yang hilang secara manual, gabungan komputer dan inspeksi manusia. Keluaran dari proses *data cleaning* adalah data yang dibersihkan secara memadai.

2. *Data Integration*

Data Integration adalah proses di mana data dari berbagai sumber data diintegrasikan menjadi satu. Data terletak dalam format berbeda di lokasi berbeda. Data dapat disimpan dalam *database*, *file teks*, *spreadsheet*, dokumen, kubus data, Internet dan sebagainya. *Data Integration* adalah tugas yang sangat kompleks dan rumit karena data dari sumber yang berbeda biasanya tidak cocok. Misalkan tabel A berisi entitas bernama *customer_id* sedangkan tabel lain B berisi entitas bernama nomor. Sangat sulit untuk memastikan apakah kedua entitas ini merujuk pada nilai yang sama atau tidak. *Metadata* dapat digunakan secara efektif untuk mengurangi kesalahan dalam proses integrasi data. Masalah lain yang dihadapi adalah redundansi data. Data yang sama mungkin tersedia di tabel berbeda dalam *database* yang sama atau bahkan di sumber data berbeda. *Data Integration* mencoba untuk mengurangi redundansi ke tingkat semaksimal mungkin tanpa memengaruhi keandalan data.

3. *Data Selection*

Proses *data mining* membutuhkan volume data historis yang besar untuk dianalisis. Jadi, biasanya *repositori* data dengan data terintegrasi berisi lebih banyak data daripada yang sebenarnya dibutuhkan. Dari data yang tersedia, data yang diminati perlu diseleksi dan disimpan. *Data Selection* adalah proses di mana data yang relevan dengan analisis diambil dari *database*.

4. *Data Transformation*

Data Transformation adalah proses mengubah dan mengkonsolidasikan data ke dalam berbagai bentuk yang cocok untuk penambangan. Transformasi data biasanya melibatkan normalisasi, agregasi, generalisasi, dll. Misalnya, kumpulan data yang tersedia sebagai "-5, 37, 100, 89, 78" dapat diubah menjadi "-0.05, 0.37, 1.00, 0.89, 0.78". Di sini data menjadi lebih cocok untuk *data mining*. Setelah integrasi data, data yang tersedia siap untuk penambangan data.

5. *Data Mining*

Data mining adalah proses inti di mana sejumlah metode kompleks dan cerdas diterapkan untuk mengekstrak pola dari data. Proses *data mining* mencakup sejumlah tugas seperti asosiasi, klasifikasi, prediksi, pengelompokan, analisis deret waktu dan sebagainya.

6. *Pattern Evaluation*

Pattern evaluation mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan berbagai jenis ukuran ketertarikan. Suatu pola dianggap menarik jika berpotensi berguna, mudah dipahami oleh manusia, memvalidasi beberapa hipotesis yang ingin dikonfirmasi atau divalidasi pada data baru dengan derajat kepastian tertentu.

7. *Knowledge Presentation*

Tahap di mana teknik-teknik visualisasi dan representasi pengetahuan digunakan untuk menghasilkan pengetahuan yang dibutuhkan

2.1.1 Pengelompokan *Data Mining*

Menurut Swastina (2013:4), *Data Mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu :

1. Deskripsi

Terkadang penelitian secara sederhana bertujuan untuk mencoba mencari cara dengan menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi bertujuan untuk mengidentifikasi pola yang muncul berulang kali pada suatu data dan mengubah pola tersebut menjadi aturan dan kriteria yang dapat dengan mudah dipahami oleh para ahli di domain aplikasi. Aturan yang dihasilkan harus mudah dipahami agar dapat meningkatkan tingkat pengetahuan dalam sistem secara efektif.

2. Prediksi

Tugas prediksi memprediksi kemungkinan nilai data yang hilang atau memprediksi masa mendatang. Prediksi melibatkan pengembangan model berdasarkan data yang tersedia dan model ini digunakan dalam memprediksi nilai mendatang dari kumpulan data baru yang menarik. Misalnya, model dapat memprediksi pendapatan karyawan berdasarkan pendidikan, pengalaman, dan faktor demografis lain seperti tempat tinggal, jenis kelamin dan lainnya. Analisis prediksi juga digunakan di berbagai bidang.

3. Estimasi

Estimasi hampir sama dengan prediksi, hanya saja variabel target diperkirakan lebih ke arah numerik daripada arah kategori. Model dibangun dengan menggunakan catatan pelengkap yang memberikan nilai variabel target sebagai nilai prediksi. Selanjutnya pada *review* selanjutnya nilai taksiran variabel target didasarkan pada nilai variabel prediktif. Misalnya, perkiraan tekanan darah sistolik pasien rumah sakit didasarkan pada usia pasien, jenis kelamin, berat badan, dan kadar natrium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediktif dalam proses pembelajaran akan menghasilkan model estimasi.

4. Klasifikasi

Dalam klasifikasi terdapat target variabel kategori. Contoh klasifikasi pendapatan dapat dibedakan dalam tiga variabel kategori yaitu: pendapatan rendah, sedang dan tinggi. Klasifikasi mendapatkan model untuk menentukan kelas suatu objek berdasarkan atributnya. Kumpulan catatan akan tersedia, setiap catatan dengan sekumpulan atribut. Salah satu atributnya adalah atribut kelas dan tujuan dari tugas klasifikasi adalah menetapkan atribut kelas ke kumpulan *record* baru seakurat mungkin. Klasifikasi dapat digunakan dalam *direct marketing*, yaitu mengurangi biaya pemasaran dengan menargetkan sekumpulan pelanggan yang cenderung membeli produk baru. Menggunakan data yang tersedia, dimungkinkan untuk mengetahui pelanggan mana yang membeli produk serupa dan siapa yang tidak membeli di masa lalu. Oleh karena itu, keputusan beli, jangan beli membentuk atribut kelas dalam kasus ini. Setelah atribut kelas ditetapkan, informasi demografis dan gaya hidup pelanggan yang membeli produk serupa dapat dikumpulkan dan surat promosi dapat dikirim langsung kepada mereka.

5. Pengklasteran

Pengklusteran digunakan untuk mengidentifikasi objek data yang mirip satu sama lain. Kesamaan tersebut dapat ditentukan berdasarkan sejumlah faktor seperti perilaku pembelian, daya tanggap terhadap tindakan tertentu, lokasi geografis dan sebagainya. Misalnya, perusahaan asuransi dapat mengelompokkan pelanggannya berdasarkan usia, tempat tinggal dan pendapatan. Informasi grup ini akan membantu untuk memahami pelanggan dan karenanya memberikan layanan yang disesuaikan.

6. Asosiasi

Teknik penambangan data ini membantu menemukan hubungan antara dua atau lebih item. Asosiasi menemukan koneksi di antara sekumpulan item. Asosiasi mengidentifikasi hubungan antar objek. Analisis asosiasi digunakan untuk manajemen komoditas, periklanan, desain katalog, pemasaran langsung dan lain-lain. Pengecer dapat mengidentifikasi produk yang biasanya dibeli oleh pelanggan

bersama-sama atau bahkan menemukan pelanggan yang menanggapi promosi jenis produk yang sama. Jika pengecer menemukan bahwa air mineral dan popok kebanyakan dibeli bersama-sama, maka dapat menjual popok untuk mempromosikan penjualan air mineral.

2.2 Machine Learning

Metode klasifikasi *machine learning* telah lama digunakan di bidang *data mining* dan banyak bidang lain dari ilmu komputer sebagai algoritma klasifikasi yang melibatkan metode dari statistik, kecerdasan buatan dan manajemen *database*, juga dapat dikategorikan sebagai kunci elemen dalam interpretasi data dan visualisasi data (Berry & Linoff, 2011). *Machine learning* membutuhkan data untuk belajar sehingga juga diistilahkan dengan *learn from data* (Alpaydin, 2010).

Machine learning adalah kumpulan metode, prinsip, dan algoritme yang memungkinkan pembelajaran dan prediksi berdasarkan data masa lalu. *Machine learning* digunakan untuk membangun model baru dan mencari model terbaik yang cocok dengan data pengujian. Metode *Machine learning* biasanya menggunakan heuristik saat mencari model. Data mining menggunakan sejumlah metode *Machine learning* termasuk pembelajaran konsep induktif, pengelompokan konseptual, dan induksi pohon keputusan.

2.3 Decision Tree

Decision tree atau pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan terkenal, selain itu juga berguna untuk mengeksplorasi data dan menemukan hubungan yang tersembunyi (Witten & Frank, 2016). Menurut Saputra (2014), *decision tree* merupakan metode yang paling efisien untuk menyaring sesuatu lewat pohon keputusan apakah suatu data lolos atau tidak terhadap saringan dengan proses yang cukup cepat dengan tahapan dalam membuat sebuah pohon keputusan sebagai berikut:

1. Menyiapkan *data training* yang sudah dikelompokkan ke dalam kelas tertentu.

2. Menentukan akar dari atribut, di mana akar akan diambil dari atribut yang terpilih yaitu dengan cara menghitung nilai gain dari masing-masing atribut. Nilai gain yang tertinggi akan menjadi akar pertama

2.4 Pengertian *Random Forest (RF)*

Istilah *Random Forest* diusulkan pertama kali oleh Tin Kam Ho pada tahun 1995, lalu dikembangkan oleh Leo Breiman pada tahun 2001. *Random Forest* merupakan salah satu metode *ensemble* untuk meningkatkan akurasi suatu klasifikasi data dari sebuah pemilah tunggal yang tidak stabil melalui kombinasi banyak pemilah dari suatu metode yang sama dengan proses voting untuk memperoleh prediksi klasifikasi akhir (Wezel & Potharst, 2007).

Metode *random forest* adalah pengembangan dari metode *Classification and Regression Tree (CART)*, yaitu dengan menerapkan metode *bootstrap aggregating (bagging)* dan *random feature selection*. *CART* sendiri merupakan metode eksplorasi data yang didasarkan pada teknik pohon keputusan di mana pohon klasifikasi dihasilkan saat peubah respon berupa data kategorik sedangkan pohon regresi dihasilkan saat peubah respons berupa data numerik. Pembangunan pohon klasifikasi *CART* meliputi tiga hal, yaitu (Breiman, 2017):

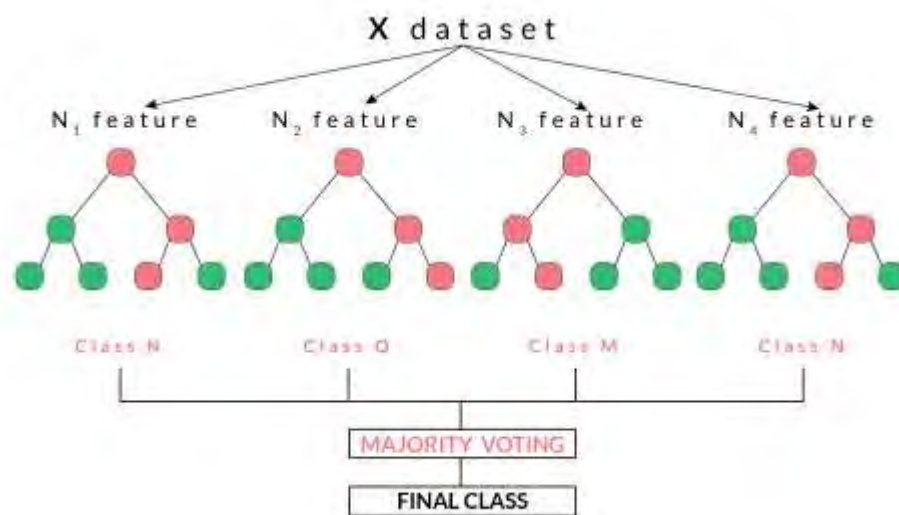
1. Pemilihan pemilah (split).
2. Penentuan simpul terminal.
3. Penandaan label kelas.

Bootstrap aggregating (Bagging) adalah teknik yang dapat digunakan untuk membentuk sampel *bootstrap* di mana setiap pohon keputusan dibangun dengan menggunakan sampel *bootstrap* dari data kandidat atribut untuk dibagi di setiap *node* yang berasal dari himpunan atribut acak dari hasil data yang dipilih dan disimpan (Mambang & Byna, 2017). *Bagging* dan *Boosting* merupakan metode *ensemble* yang relatif baru namun telah menjadi populer (Nidhomuddin, 2015).

Random forest adalah klasifikasi yang terdiri dari beberapa pohon keputusan yang dibangun dengan menggunakan vektor acak (Mambang & Byna, 2017). Menurut Adnyana (2015), *random forest* merupakan pengembangan dari *decision tree* dengan

menggunakan beberapa *decision tree* di mana setiap *decision tree* telah dilakukan *training* menggunakan sampel individu dan setiap atribut dipecah pada *tree* yang dipilih antara atribut yang bersifat acak dan dalam proses klasifikasi dilakukan *vote* dari suara terbanyak di populasi *tree*. *Random forest* melalui proses pengacakan yang tidak hanya dilakukan pada data sampel saja melainkan juga pada pengambilan variabel bebas sehingga pohon klasifikasi yang dibangkitkan akan memiliki ukuran dan bentuk yang berbeda-beda.

Representasi data dalam bentuk pohon memiliki keunggulan dibandingkan dengan pendekatan lain yang berarti dan mudah ditafsirkan. Tujuannya adalah untuk membuat model klasifikasi yang memprediksi nilai atribut target berdasarkan beberapa atribut masukan dari sampel *dataset*.



Sumber: <https://www.quantinsti.com>

Gambar 2.2 *Random Forest*

Dalam gambar 2.2 operator *random forest* menghasilkan satu set pohon acak, kelas yang dihasilkan dari proses klasifikasi dipilih dari kelas yang paling banyak (modus) yang dihasilkan oleh pohon acak yang ada (Biau, 2012). Banyak pohon ditumbuhkan dalam metode *random forest*, sehingga terbentuk hutan (*forest*) yang akan dianalisis. Pada gugus data yang terdiri atas n amatan dan p peubah penjelas, *random forest* dilakukan dengan cara (Breiman, 2001):

1. Melakukan penarikan contoh acak berukuran n dengan pemulihan pada gugus data di mana tahapan ini merupakan tahapan *bootstrap*.
2. Dengan menggunakan contoh bootstrap, pohon dibangun sampai mencapai ukuran maksimum (tanpa pemangkasan). Pada setiap simpul, pemilihan pemilah dilakukan dengan memilih m peubah penjelas secara acak, di mana $m \ll p$, lalu pemilah terbaik dipilih berdasarkan m peubah penjelas tersebut di mana tahapan ini disebut dengan tahapan *random feature selection*.
3. Ulangi langkah 1 dan 2 sebanyak k kali, sehingga terbentuk sebuah hutan yang terdiri atas k pohon

2.5 RapidMiner Studio

Rapid miner studio adalah sebuah solusi untuk melakukan analisis terhadap *data mining*, *text mining* dan analisis prediksi. Untuk membuat keputusan yang baik *Rapid miner* memberikan wawasan kepada pengguna dengan teknik deskriptif dan prediksi. terdapat 500 operator *data mining*, termasuk operator untuk *data preprocessing*, visualisasi, *input* dan *output*. *Rapid miner* merupakan *software* yang tidak tergantung pada *software* lain untuk analisis data dan mesin *data mining*.



Sumber: <https://rapidminer.com>

Gambar 2.3 Jendela Rapid miner

Pada gambar 2.3 merupakan jendela editor *Rapidminer* pada nomor 1 merupakan *Operator* berfungsi untuk Blok penyusun yang digunakan untuk membuat proses *Rapid miner*, nomor 2 *Repository* berfungsi untuk Penyimpanan dalam *Rapid miner studio* untuk data dan proses *Rapid miner*. Nomor 3 *Process panel (Main Process)* berfungsi untuk area kerja untuk proses pembangunan. Nomor 4 *Views* berfungsi untuk area kerja untuk mengakses fungsionalitas tertentu. Nomor 5 *Ports* berfungsi untuk mekanisme *input* dan *output* untuk operator dan proses. Nomor 6 *Parameters* berfungsi untuk pengaturan yang mengubah perilaku operator. Terakhir nomor 7 *Help* berfungsi untuk bantuan peka konteks untuk operator terpilih.

2.6 Lalu Lintas

Lalu lintas merupakan pergerakan orang dan kendaraan di ruang lalu lintas jalan. Ruang Lalu Lintas Jalan merupakan tempat atau prasarana yang untuk tempat gerak atau perpindahan kendaraan, orang, dan barang yang berupa Jalan beserta fasilitas pendukung lainnya. Kepadatan lalu lintas adalah ukuran atau volume kendaraan yang melewati jalan di daerah tertentu dengan arus kendaraan yang bervariasi di saat jam-jam tertentu di nyatakan dalam per jam per kilometer. Menurut aturan pemerintah pada undang-undang tahun 2009 Nomor 22 tentang Lalu lintas dan angkutan jalan. Menurut Pasal 1 Undang- undang Nomor 22 tahun 2009, lalu lintas didefinisikan sebagai gerak kendaraan dan orang di ruang lalu lintas jalan. Ruang lalu lintas jalan didefinisikan prasarana yang diperuntukkan bagi gerak pindah kendaraan, orang, dan/atau barang yang berupa jalan dengan fasilitas pendukungnya.

Kinerja jalan berfungsi untuk mengetahui kemampuan jalan di dalam menampung kendaraan. Dengan kata lain bahwa untuk mengetahui apakah jalan tersebut sudah mencapai kapasitasnya atau belum apabila sudah mencapai kapasitasnya maka penanganan biasa dilakukan *stakeholder* adalah dengan melakukan penambahan lajur atau pembangunan jalan baru. Tingkat pelayanan jalan dapat diukur dengan menggunakan arus lalu lintas dan waktu tempuh, kapasitas jalan, volume jalan, *Volume Capacity Ratio* dan *Level of Service*.

2.6.1 Volume Capacity Ration

Volume capacity ratio merupakan perbandingan antara volume yang melintas (smp) dengan kapasitas pada suatu ruas jalan tertentu (smp). Volume lalu lintas diperoleh berdasarkan survei yang dilakukan, sedangkan nilai kapasitas diperoleh dari lingkungan ruas jalan dan survei geometrik yang meliputi potongan melintang, persimpangan, *alinyamen* vertikal dan *alinyamen* horizontal. Selanjutnya dihitung berdasarkan model yang di kembangkan oleh MKJI atau Manual Kapasitas Jalan Indonesia 1997. Adapun tingkat pelayanan (VCR) dilakukan dengan persamaan sebagai berikut :

$$VCR = V/C \quad (2.1)$$

Keterangan :

VCR : Volume kapasitas ratio (nilai tingkat pelayanan)

V : Volume lalu lintas (smp/jam)

C : Kapasitas ruas jalan (smp/jam)

Sedangkan standar nilai VCR ditetapkan berdasarkan MKJI (Manual Kapasitas Jalan Indonesia) adalah sebagai berikut :

Tabel 2.1 Kriteria Tingkat Pelayanan

	Tingkat Pelayanan	Nilai
A	Sangat Tinggi	0,00-0,20
B	Tinggi	0,21-0,44
C	Sedang	0,45-0,74
D	Rendah	0,75-0,84
E	Sangat Rendah	0,85-1,00
F	Sangat Sangat Rendah	>1,00

Sumber : MKJI 1997

2.6.2 Kapasitas Jalan

Kapasitas jalan merupakan daya tampung suatu jalan untuk jumlah kendaraan maksimum yang dapat bergerak dalam periode waktu tertentu. Kapasitas jalan biasanya dinyatakan dalam Satuan Mobil Penumpang (smp) per jam

Faktor-faktor yang mempengaruhi kapasitas jalan adalah lebar jalur atau lajur, ada tidaknya pemisah/median jalan, hambatan bahu jalan, lokasi didaerah perkotaan atau lokasi jalan luar kota. Persamaan untuk menghitung kapasitas jalan adalah sebagai berikut :

$$C = C_0 \times FC_w \times FC_{sp} \times FC_{sf} \times FC_{cs} \quad (2.2)$$

Keterangan :

- C : Kapasitas (smp/jam)
 C₀ : Kapasitas dasar (smp/jam)
 FC_w : Faktor penyesuaian lebar jalan 20
 FC_{sp} : Faktor penyesuaian pemisah arah (hanya untuk jalan tak terbagi)
 FC_{sf} : Faktor penyesuaian hambatan samping dan bahu jalan
 FC_{cs} : Faktor penyesuaian ukuran kota

2.6.3 Volume Lalu Lintas

Banyaknya kendaraan yang melewati suatu titik atau garis tertentu pada suatu penampang melintang jalan disebut Volume lalu lintas. Hasil pencacahan volume lalu lintas merupakan informasi yang diperlukan untuk fase desain, perencanaan, manajemen sampai pengoperasian jalan. Jenis kendaraan diklasifikasikan dalam 3 (tiga) macam kendaraan yaitu:

1. *Light Vehicles (LV)* / Kendaraan Ringan

LV merupakan pengenal untuk kendaraan bermotor mobil penumpang dengan roda empat

2. *Heavy Vehicles (HV)* / Kendaraan berat

HV merupakan pengenal untuk kendaraan bermotor dengan roda lebih dari empat seperti Bus, truk dua gandar, truk tiga gandar dan kombinasi yang sesuai

3. *Motor Cycle (MC)* / Sepeda motor

MC merupakan pengenal kendaraan bermotor untuk roda dua.

Becak , sepeda, dan kereta dorong yang parkir pada badan jalan dan pejalan kaki dianggap sebagai hambatan samping. Total kendaraan yang lewat lalu dihitung

dengan satuan mobil penumpang (smp) perjam untuk setiap kendaraan, dengan faktor koreksinya masing-masing. Total arus lalu lintas dalam smp/jam adalah:

$$Q = (\text{empLV} \times \text{LV} + \text{empHV} \times \text{HV} + \text{empMC} \times \text{MC}) \quad (2.3)$$

Keterangan:

Q = volume kendaraan bermotor (smp/jam)

EmpLV = nilai ekivalen mobil penumpang untuk kendaraan ringan

EmpHV = nilai ekivalen mobil penumpang untuk kendaraan berat

EmpMC = nilai ekivalen mobil penumpang untuk sepeda motor

LV = notasi untuk kendaraan ringan

HV = notasi untuk kendaraan berat

MC = notasi untuk sepeda motor

Tabel 2.2 Kriteria Tingkat Pelayanan

Tipe Kendaraan	Nilai Ekivalen
Kendaraan Ringan (LV)	1,0
Kendaraan Berat (HV)	1,3
Sepeda Motor (MC)	0,5

Sumber: MKJI 1997

Yang nantinya hasil faktor satuan mobil penumpang (P) ini dimasukkan dalam rumus volume lalu lintas:

$$Q = P \times Q_v \quad (2.4)$$

Dengan:

Q = Volume kendaraan bermotor (smp/jam),

P = Faktor satuan mobil penumpang,

Q_v = Volume kendaraan bermotor (kendaraan per jam)

2.7 Penelitian Terkait

Dalam penelitian ini, penulis sedikit banyak terinspirasi dan mereferensi dari penelitian-penelitian sebelumnya yang berkaitan dengan latar belakang masalah pada penelitian ini. Dijelaskan dalam Tabel 2.3 adapun penelitian yang berhubungan dengan penelitian ini antara lain yaitu: Penelitian yang dilakukan oleh Tin Kam Ho ,1998 yang

berjudul “*The Random Subspace Method for Constructing Decision Forests*”. Penelitian ini merupakan penelitian tentang *Random Forest* yang di terbitkan oleh IEEE *Transactions on Pattern Analysis and Machine Intelligence*. Penelitian selanjutnya dilakukan oleh Hikmat Iskandar, yang berjudul “Volume Lalu Lintas Untuk Geometrik Dan Perkerasan Jalan”, Penelitian ini menjelaskan bagaimana data volume lalu lintas digunakan untuk perkerasan jalan. Selanjutnya penelitian dilakukan oleh Disi M Hanafiah, Handyana Arifin dan Agus Bari yang berjudul “*Optimization Of Image Processing Technology (Ipt) For Counting Motorcycles In Indonesia*”, Penelitian ini fokus ke optimalisasi teknologi IPT untuk menghitung kendaraan. Kemudian penelitian dilakukan oleh Disi M Hanafiah, yang berjudul “Akurasi Alat Penghitung Lalu Lintas Plato 2.1 Berbasis Pengolahan Citra - Background Substraction”, Penelitian ini menjelaskan tentang teknologi *Traffic Counting* yang telah dirancang beserta akurasi. Hasil dari alat ini merupakan sumber *Datasheet*. Berdasarkan penelitian yang dilakukan oleh Antisa Kurnia Hayatri mengenai penelitian yang berjudul “Perbandingan Metode *Random Forest* Dan Algoritma C4.5 Pada Diagnosis Penyakit Liver”, Antisa menyimpulkan dari hasil penelitian Akurasi *Random Forest* lebih baik dari pada C4.5. Berikut tabel di bawah ini merupakan penelitian terkait:

Tabel 2.3 Perbandingan Penelitian Terkait

No	Judul Penelitian	Instansi	Peneliti	Keterangan
1	The Random Subspace Method for Constructing Decision Forests	IEEE	Tin Kam Ho	Penelitian tentang <i>Random Forest</i>
2	Volume Lalu Lintas Untuk Geometrik Dan Perkerasan Jalan	Puslitbang Jalan Dan Jembatan	Hikmat Iskandar	Penelitian Lalu lintas fokus ke pengerasan jalan
3	Optimization Of Image Processing Technology (IPT) For Counting Motorcycles In Indonesia	Puslitbang Jalan Dan Jembatan	Disi M Hanafiah, Handyana Arifin, Agus Bari	Fokus ke optimalisasi IPT untuk menghitung kendaraan
4	Akurasi Alat Penghitung Lalu Lintas Plato 2.1 Berbasis Pengolahan Citra - Background Substraction	Puslitbang Jalan Dan Jembatan	Disi M Hanafiah	Alat sumber datasheet yang digunakan dalam penelitian

No	Judul Penelitian	Instansi	Peneliti	Keterangan
5	Klasifikasi Kepadatan Lalu Lintas Berbasis <i>Multitracking Object</i> Dengan menggunakan metode Particle Filter	Universitas Telkom	Hafizd AIDjohari	Fokus penelitian ke klasifikasi Image Processing
6	Analisa Dan Implementasi Algoritma RF sebagai sebuah Clasifier dalam <i>Data Mining</i>	Universitas Telkom	Rahmi Fitriani	Penelitian lebih ke membangun perangkat lunak
7	Klasifikasi Data Time Series Arus Lalu Lintas Jangka Pendek Menggunakan Algoritma <i>Adabost</i> Dengan <i>Random Forest</i>	Universitas Dian Nuswantoro	Ahmad Rofiqul, Musliqh, Heru Agus Santoso, Aris Marjuni	Datasheet Minim Hanya 7 baris data diambil dari Binamarga Jatim, Jenis Datasheet yang dipakai berbeda, Fokus ke Algoritma <i>Adabost</i>
8	Klasifikasi Rekomendasi Dosen Pembimbing Tugas Akhir Dengan Menggunakan Algoritma <i>Random Forest (RF)</i>	Universitas Muhammadiyah Malang	Rosa Regina	Tentang klasifikasi rekomendasi dosen pembimbing
9	Klasifikasi Penyakit Diabetes Mellitus Dengan Menggunakan Perbandingan Algoritma J48 Dan Random Forest (Studi Kasus : Rumah Sakit Muhammadiyah Lamongan)	Universitas Muhammadiyah Malang	Shaumi Rizqi Maziida	Dalam Penelitian ini akurasi <i>Random Forest</i> lebih baik dibandingkan J48
10	Perbandingan Metode Random Forest Dan Algoritma C4.5 Pada Diagnosis Penyakit Liver	Ull	Antisa Kurnia Hayatri	Akurasi <i>Random Forest</i> lebih baik dari pada C4.5
11	Klasifikasi Ketepatan Lama Studi Mahasiswa Menggunakan Metode Support Vector Machine Dan Random Forest (Studi Kasus : Data Lama Studi Alumni Universitas Islam Indonesia Tahun Kelulusan 2000-2017)	Ull	Ahmad Rofiqul, Musliqh, Heru Agus Santoso, Aris Marjuni	Akurasi <i>Random Forest</i> lebih baik dari pada SVM
12	Prediksi Intensitas Hujan Kota Surabaya dengan MatLab menggunakan Teknik Random Forest dan CART (Studi Kasus Kota Surabaya)	Institut Teknologi Sepuluh November	Maulana Dhawangkhara	<i>Random Fores</i> memiliki Ferforma lebih baik dibandingkan CART

BAB III

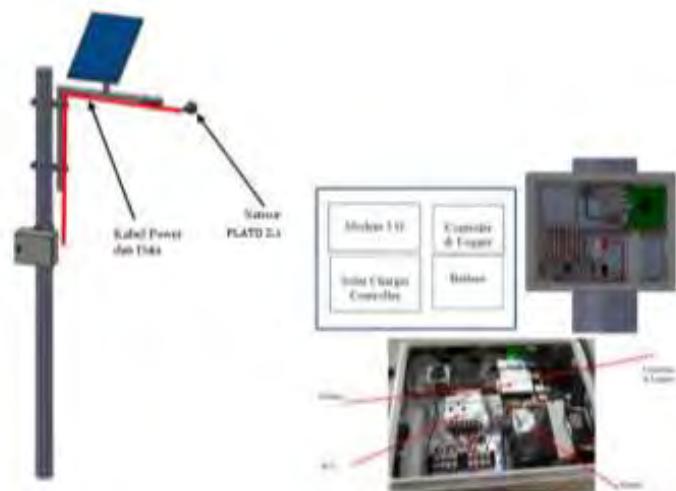
OBJEK DAN METODE PENELITIAN

3.1 Kepadatan Lalu Lintas

Objek penelitian dalam penelitian ini merupakan data hasil hitungan jumlah kendaraan, data tersebut diambil dari sebuah alat pencacah lalu lintas elektronik, di mana data kendaraan yang lewat pada alat dapat membedakan menjadi 5 (lima) buah golongan berdasarkan dimensi:

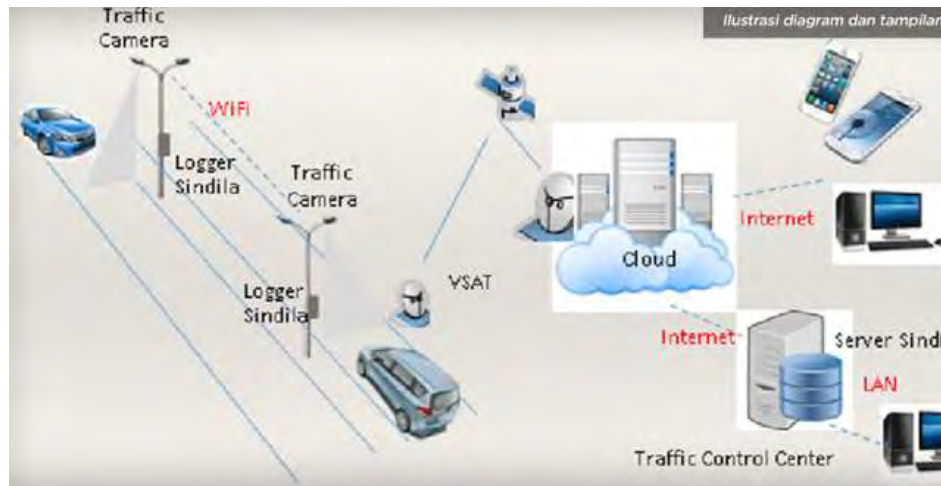
1. Golongan I: Sepeda motor
2. Golongan II: Kendaraan ringan
3. Golongan III: Kendaraan berat menengah
4. Golongan IV: Bus besar
5. Golongan V : Truk besar

Alat pencacah lalu lintas elektronik (*Traffic Counting*) ini merupakan hasil penelitian Pusat Litbang Jalan dan Jembatan yang bekerja sama dengan PT Industri Telekomunikasi Indonesia untuk diproduksi dan diperjualbelikan dalam kualitas dan skala industri.



Sumber: Puslitbang jalan dan jembatan (2015)

Gambar 3.1 Komponen Sistem



Sumber: <http://www.pusjatan.pu.go.id>

Gambar 3.2 Ilustrasi Cara Kerja Sistem

Gambar 3.2 menggambarkan cara kerja umum sistem ini setiap kendaraan yang lewat akan dihitung dan disimpan dalam *logger* lalu dikirim secara berkala ke server menggunakan koneksi internet. Data yang sudah tersimpan banyak dari server ini merupakan *dataseet* untuk penelitian.



Sumber: www.beritasatu.com

Gambar 3.3 Ilustrasi Lalu lintas di jalan raya By Pass Jomin

Pada gambar 3.3 menggambarkan arus lalu lintas di sekitar simpang Jomin hingga ke jalur Pantura wilayah Kabupaten Karawang, Jawa Barat, cukup padat pada hari libur Natal dan libur sekolah. Kendaraan yang melintas di sekitar simpang Jomin itu didominasi kendaraan pribadi. Sedangkan kendaraan jenis bus dan truk hanya sebagian

kecil yang melintas. Selain akibat tingginya volume kendaraan yang melintas, arus lalu lintas sekitar simpang Jomin padat karena terjadi pertemuan arus serta terjadi penyempitan ruas jalan. (Berita satu, 2014)

3.2 Metode Penelitian

Metode penelitian merupakan suatu tahapan atau langkah secara ilmiah untuk mengumpulkan data kemudian menjadi sebuah kesimpulan ilmiah untuk mencapai tujuan dari penelitian. Tahapan proses yang akan dilakukan dalam penelitian ini digambarkan dalam diagram alir sebagai berikut:



Gambar 3.4 Diagram Alir Penelitian

Gambar diagram alir penelitian yang digambarkan pada gambar 3.4 dimulai dari tahap awal dalam penelitian ini dimulai dari survei pendahuluan dengan wawancara pihak yang berkompeten dalam lalu lintas di antaranya Dinas Perhubungan dan Pusat Penelitian dan Pengembangan Jalan dan Jembatan, setelah mendapatkan beberapa permasalahan yang mengidentifikasi mulai menentukan topik penelitian, menentukan rumusan permasalahan dan tujuan penelitian, untuk penelitian agar tidak meluas maka ditentukan ruang lingkup penelitian dan menentukan lokasi data lalu lintas yang akan dijadikan bahan penelitian. Tahapan berikutnya studi *literatur* dengan banyak membaca jurnal, teori dalam buku dan studi beberapa kasus.

3.2.1 Pengumpulan Data

Untuk mendapatkan *dataset* yang diperlukan maka digunakan beberapa metode pengumpulan data. *Dataset* diambil dari sebuah server alat pencacah lalu lintas elektronik dengan lokasi perangkat yang terpasang di ruas jalan *By Pass* Jomin Cikampek Karawang Jawa Barat.

3.2.2 Metode Analisis

Data yang didapat masih perlu dilakukan normalisasi terlebih dahulu. Tahapan normalisasi dalam *data mining* dilakukan pada fase *Pre-processing*. Dalam tahapan ini dilakukan pengkajian data yang telah diperoleh berdasarkan teori, yaitu analisis data secara deskriptif dan menggunakan metode *random forest*, kemudian dilakukan penerapan langkah-langkah sebagai berikut:

1. Menentukan data latih (*training*) yang akan digunakan untuk melakukan metode *random forest*.
2. Menentukan nilai *Mtry* terbaik.
3. Memilih nilai *n* yang menunjukkan jumlah pohon.
4. Membuat plot untuk melihat proses pelatihan data latih agar mendapatkan hasil yang lebih optimal.
5. Memilih teknik sampling yang akan digunakan pada metode *random forest*.

6. Mengklasifikasi data dengan metode *random forest* menggunakan data latih (*training*) yang telah ditentukan sebelumnya.
7. Melakukan pengujian akurasi menggunakan data uji (*testing*) untuk melihat seberapa besar akurasi metode *random forest* terhadap data penelitian.

3.3 Persiapan Data (*Data Preparation*)

Data Preparation atau bisa disebut juga dengan *data preprocessing* adalah suatu proses atau langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas. Adapun data lalu lintas yang telah didapatkan dalam bentuk *file* XLSX. Data berupa jumlah kendaraan melintas setiap hari dari tanggal 1 Maret 2018 sampai 30 Juni 2018 berjumlah 34.560 data diperoleh dari server PT Inti Konten Indonesia yang merupakan anak perusahaan dari PT Industri Telekomunikasi Indonesia (Persero) dan sudah diizinkan oleh peneliti Pusjatan sebagai pemilik lisensi dari perangkat pencacah lalu lintas elektronik. Berikut struktur tabel data lalu lintas tersebut:

Tabel 3.1 Struktur Tabel Data Lalu Lintas

No	Field	Type	Keterangan
1	Id	<i>Int(8)</i>	Merupakan nomor urut
2	Lokasi	<i>Varchar (50)</i>	Berisi Lokasi perangkat terpasang
3	Waktu	<i>Timestamp</i>	Berisi waktu pengambilan data kendaraan yang lewat selama 5 menit
4	Jenis Jalan	<i>Varchar (20)</i>	Berisi Jenis jalan, misalkan dalam kota atau luar kota
5	Total kendaraan gol I arah menjauh	<i>Int(5)</i>	Berisi total golongan I arah menjauh atau arah ke Purwakarta dalam 5 menit
6	Total kendaraan gol II arah menjauh	<i>Int(5)</i>	Berisi total golongan II arah menjauh atau arah ke Purwakarta dalam 5 menit
7	Total kendaraan gol III arah menjauh	<i>Int(5)</i>	Berisi total golongan III arah menjauh atau arah ke Purwakarta dalam 5 menit
8	Total kendaraan gol IV arah menjauh	<i>Int(5)</i>	Berisi total golongan IV arah menjauh atau arah ke Purwakarta dalam 5 menit
9	Total kendaraan gol V arah menjauh	<i>Int(5)</i>	Berisi total golongan V arah menjauh / arah ke Purwakarta dalam 5 menit
10	Kecepatan rata-rata kendaraan gol I arah menjauh	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan I arah menjauh atau arah ke Purwakarta dalam 5 menit

No	Field	Type	Keterangan
11	Kecepatan rata-rata kendaraan gol II arah menjauh	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan II arah menjauh atau arah ke Purwakarta dalam 5 menit
12	Kecepatan rata-rata kendaraan gol III arah menjauh	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan III arah menjauh atau arah ke Purwakarta dalam 5 menit
13	Kecepatan rata-rata kendaraan gol IV arah menjauh	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan IV arah menjauh atau arah ke Purwakarta dalam 5 menit
14	Kecepatan rata-rata kendaraan gol V arah menjauh	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan V arah menjauh atau arah ke Purwakarta dalam 5 menit
15	Total kendaraan gol I arah mendekat	<i>Int(5)</i>	Berisi total golongan I arah mendekat atau arah ke Cirebon dalam 5 menit
16	Total kendaraan gol II arah mendekat	<i>Int(5)</i>	Berisi total golongan II arah mendekat atau arah ke Cirebon dalam 5 menit
17	Total kendaraan gol III arah mendekat	<i>Int(5)</i>	Berisi total golongan III arah mendekat atau arah ke Cirebon dalam 5 menit
18	Total kendaraan gol IV arah mendekat	<i>Int(5)</i>	Berisi total golongan IV arah mendekat atau arah ke Cirebon dalam 5 menit
19	Total kendaraan gol V arah mendekat	<i>Int(5)</i>	Berisi total golongan V arah mendekat atau arah ke Cirebon dalam 5 menit
20	Kecepatan rata-rata kendaraan gol I arah mendekat	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan I arah mendekat atau arah ke Cirebon dalam 5 menit
21	Kecepatan rata-rata gol I arah mendekat	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan II arah mendekat atau arah ke Cirebon dalam 5 menit
22	Kecepatan rata-rata kendaraan gol III arah mendekat	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan III arah mendekat atau arah ke Cirebon dalam 5 menit
23	Kecepatan rata-rata kendaraan gol IV arah mendekat	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan IV arah mendekat atau arah ke Cirebon dalam 5 menit
24	Kecepatan rata-rata kendaraan gol V arah mendekat	<i>Int(5)</i>	Berisi kecepatan rata-rata golongan V arah mendekat atau arah ke Cirebon dalam 5 menit

Dapat diperhatikan dalam tabel di atas ada dua arah yang diambil data lalu lintasnya pertama arah menjauh dari perangkat atau arah ke Purwakarta dan arah mendekat dari perangkat atau arah ke Cirebon, pengambilan data lalu di atur dalam per lima menit, pada tabel 3.1 ada data kecepatan rata-rata kendaraan dalam penelitian ini data kecepatan tidak akan digunakan karena akan berfokus ke data jumlah kendaraan yang dilayani oleh jalan atau tingkat kepadatannya.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1 Data Cleaning

Data cleaning pada data lalu lintas ini mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, data kosong (NA), memperbaiki kesalahan dan memperkaya data yang sudah ada. Berikut gambaran data lalu lintas pada excel:

Tabel 4.1 Data Lalu Lintas Pada Excel

id	Lokasi	Timestamp	Arah Jalan	Arah Menjauh										Arah Mendekat										
				Jumlah Kendaraan					Kecepatan Rata-Rata					Jumlah Kendaraan					Kecepatan Rata-Rata (Km/h)					
				Gol 1	Gol 2	Gol 3	Gol 4	Gol 5	Gol 1	Gol 2	Gol 3	Gol 4	Gol 5	Gol 1	Gol 2	Gol 3	Gol 4	Gol 5	Gol 1	Gol 2	Gol 3	Gol 4	Gol 5	
1	J. Raya By Pass Jember	2018-03-05 00	Luar Kota	32	2	3	0	1	64	83	44	0	37	86	70	24	0	15	70	80	75	0	31	
2	J. Raya By Pass Jember	2018-03-06 00	Luar Kota	30	5	7	0	0	66	84	62	0	0	39	19	8	0	0	73	85	74	0	64	
3	J. Raya By Pass Jember	2018-03-05 00	Luar Kota	46	3	12	0	2	57	57	62	0	0	36	39	18	18	0	15	73	79	74	0	58
4	J. Raya By Pass Jember	2018-03-06 00	Luar Kota	27	7	3	0	0	62	80	58	0	0	30	7	0	0	18	62	75	78	0	44	
5	J. Raya By Pass Jember	2018-03-06 00	Luar Kota	27	4	5	0	0	64	77	77	0	0	36	11	12	0	21	72	80	81	0	58	
6	J. Raya By Pass Jember	2018-03-05 01	Luar Kota	17	10	3	0	0	59	79	62	0	0	46	18	16	0	36	70	80	72	0	52	
7	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	32	13	5	0	1	89	87	83	0	0	44	65	20	19	0	22	75	75	74	0	44
8	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	36	8	4	0	0	80	88	51	0	0	40	10	14	0	0	80	74	80	71	0	54
9	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	38	5	4	0	0	71	70	80	0	0	37	56	11	24	0	41	68	65	74	0	36
10	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	42	14	0	0	0	72	85	85	0	0	78	57	17	12	0	39	75	74	64	0	47
11	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	38	2	4	0	0	78	80	84	0	0	42	40	18	9	0	20	76	78	84	0	54
12	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	22	5	0	0	0	68	72	75	0	0	40	41	14	14	0	26	65	77	67	0	34
13	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	31	8	2	0	0	62	63	59	0	0	32	24	18	0	14	67	71	75	0	48	
14	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	57	8	7	0	0	74	79	53	0	0	44	21	18	0	19	65	71	60	0	31	
15	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	39	6	3	0	0	66	80	87	0	0	79	10	10	0	27	73	73	73	0	46	
16	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	18	3	2	0	1	63	80	80	0	0	80	58	17	11	0	27	72	78	70	0	50
17	J. Raya By Pass Jember	2018-03-06 01	Luar Kota	29	7	3	0	4	81	76	43	0	0	43	45	13	14	0	14	76	79	75	0	50
18	J. Raya By Pass Jember	2018-03-06 02	Luar Kota	37	7	0	0	0	61	67	54	0	0	45	46	19	11	0	32	76	80	65	0	42
19	J. Raya By Pass Jember	2018-03-06 02	Luar Kota	38	10	4	0	0	72	75	87	0	0	48	66	58	25	0	66	73	70	67	0	44
20	J. Raya By Pass Jember	2018-03-06 02	Luar Kota	40	4	5	0	0	83	75	72	0	0	42	71	43	17	0	37	68	67	68	0	37
21	J. Raya By Pass Jember	2018-03-06 02	Luar Kota	16	4	0	0	0	53	84	48	0	0	16	106	40	8	0	30	73	75	68	0	46

Pada tabel 4.1 menampilkan data lalu lintas dengan struktur data lalu lintas seperti yang tunjukan pada tabel 3.1 pada bab sebelumnya. Data lalu lintas pada tabel 4.1 diatas merupakan data *regular* dimana data tersebut merupakan gambaran umum kondisi lalu lintas, di mana pada saat mudik lebaran data lalu lintas tidak di ambil karena kondisi tersebut tidak menggambarkan data lalu lintas dalam setahun.

Untuk melakukan *data mining* atau penambangan data pada data lalu lintas maka dimulai dari mencari nilai volume lalu lintas berdasarkan jumlah kendaraan dengan ekuivalennya, volume lalu lintas dinyatakan dalam satuan mobil penumpang per jam (smp/jam). Selain ditambahkannya *field* baru volume lalu lintas, dalam penelitian ini ada beberapa *field* yang akan dihilangkan untuk proses *data mining*, berikut *field* yang akan dihilangkan:

1. Kecepatan rata-rata kendaraan gol I arah menjauh
2. Kecepatan rata-rata kendaraan gol II arah menjauh

3. Kecepatan rata-rata kendaraan gol III arah menjauh
4. Kecepatan rata-rata kendaraan gol IV arah menjauh
5. Kecepatan rata-rata kendaraan gol V arah menjauh
6. Kecepatan rata-rata kendaraan gol I arah mendekat
7. Kecepatan rata-rata kendaraan gol II arah mendekat
8. Kecepatan rata-rata kendaraan gol III arah mendekat
9. Kecepatan rata-rata kendaraan gol IV arah mendekat
10. Kecepatan rata-rata kendaraan gol V arah menjauh

Penghapusan beberapa *field* di atas dikarenakan dalam penelitian ini tidak memanfaatkan data kecepatan per lajur karena penelitian difokuskan ke tingkat kepadatan lalu lintas berdasarkan Volume Kapasitas Rasio (VCR).

Tabel 4.2 Struktur Tabel Setelah Data Kecepatan Di Hapus

No	Field	Type	Keterangan
1	Id	<i>Int(8)</i>	Merupakan nomor urut
2	Lokasi	<i>Varchar (50)</i>	Berisi Lokasi perangkat terpasang
3	Waktu	<i>Timestamp</i>	Berisi waktu pengambilan data kendaraan yang lewat selama 5 menit
4	Jenis Jalan	<i>Varchar (20)</i>	Berisi Jenis jalan, misalkan dalam kota atau luar kota
5	Total kendaraan gol I arah menjauh	<i>Int(5)</i>	Berisi total golongan I arah menjauh atau arah ke Purwakarta dalam 5 menit
6	Total kendaraan gol II arah menjauh	<i>Int(5)</i>	Berisi total golongan II arah menjauh atau arah ke Purwakarta dalam 5 menit
7	Total kendaraan gol III arah menjauh	<i>Int(5)</i>	Berisi total golongan III arah menjauh atau arah ke Purwakarta dalam 5 menit
8	Total kendaraan gol IV arah menjauh	<i>Int(5)</i>	Berisi total golongan IV arah menjauh atau arah ke Purwakarta dalam 5 menit
9	Total kendaraan gol V arah menjauh	<i>Int(5)</i>	Berisi total golongan V arah menjauh / arah ke Purwakarta dalam 5 menit
10	Total kendaraan gol I arah mendekat	<i>Int(5)</i>	Berisi total golongan I arah mendekat atau arah ke Cirebon dalam 5 menit
11	Total kendaraan gol II arah mendekat	<i>Int(5)</i>	Berisi total golongan II arah mendekat atau arah ke Cirebon dalam 5 menit
12	Total kendaraan gol III arah mendekat	<i>Int(5)</i>	Berisi total golongan III arah mendekat atau arah ke Cirebon dalam 5 menit
13	Total kendaraan gol IV arah mendekat	<i>Int(5)</i>	Berisi total golongan IV arah mendekat atau arah ke Cirebon dalam 5 menit

No	Field	Type	Keterangan
14	Total kendaraan gol V arah mendekat	Int(5)	Berisi total golongan V arah mendekat atau arah ke Cirebon dalam 5 menit
15	Volume lalu lintas	Int(5)	Volume lalu lintas dengan satuan smp/jam

4.3 Data Transformation

Pada tahap ini data lalu lintas akan diubah lagi menjadi bentuk yang tepat untuk menambang data dan data lalu lintas diintegrasikan dengan kapasitas jalan pada jalan *By Pass* Jomin Cikampek. Data kapasitas jalan diambil dari lokasi jalan *By Pass* Jomin Cikampek dengan mengukur lebar jalur, ada tidaknya pemisah antar jalur atau median serta nilai hambatan samping.



Gambar 4.1 Survei Jalan Untuk Mendapatkan Nilai Kapasitas Jalan

Berdasarkan pengamatan lokasi Jalan *By Pass* Jomin merupakan tipe jalan dengan 2 lajur tak terbagi (2/2 UD), jalan tanpa adanya median atau pemisah antar jalur. Jalan *By Pass* Jomin termasuk kategori jalan luar kota, arah mendekat merupakan arah

ke Cirebon dan arah menjauh merupakan arah ke Purwakarta, di lokasi ini arah ke Cirebon jauh lebih banyak dari pada arah ke Purwakarta.



Gambar 4.2 Mengamati Untuk Mendapat Nilai Hambatan Sampung

Dengan survei mendapatkan data kapasitas jalan dengan satuan mobil penumpang per jam maka struktur tabel berubah menjadi sebagai berikut:

Tabel 4.3 Struktur Tabel Setelah Mencantumkan Nilai Kapasitas Jalan

No	Field	Type	Keterangan
1	Id	Int(8)	Merupakan nomor urut
2	Lokasi	Varchar (50)	Berisi Lokasi perangkat terpasang
3	Waktu	Timestamp	Berisi waktu pengambilan data kendaraan yang lewat selama 5 menit
4	Jenis Jalan	Varchar (20)	Berisi Jenis jalan, misalkan dalam kota atau luar kota
5	Total kendaraan gol I arah menjauh	Int(5)	Berisi total golongan I arah menjauh atau arah ke Purwakarta dalam 5 menit
6	Total kendaraan gol II arah menjauh	Int(5)	Berisi total golongan II arah menjauh atau arah ke Purwakarta dalam 5 menit
7	Total kendaraan gol III arah menjauh	Int(5)	Berisi total golongan III arah menjauh atau arah ke Purwakarta dalam 5 menit
8	Total kendaraan gol IV arah menjauh	Int(5)	Berisi total golongan IV arah menjauh atau arah ke Purwakarta dalam 5 menit
9	Total kendaraan gol V arah menjauh	Int(5)	Berisi total golongan V arah menjauh / arah ke Purwakarta dalam 5 menit
10	Total kendaraan gol I arah mendekat	Int(5)	Berisi total golongan I arah mendekat atau arah ke Cirebon dalam 5 menit
11	Total kendaraan gol II arah mendekat	Int(5)	Berisi total golongan II arah mendekat atau arah ke Cirebon dalam 5 menit

No	Field	Type	Keterangan
12	Total kendaraan gol III arah mendekat	<i>Int(5)</i>	Berisi total golongan III arah mendekat atau arah ke Cirebon dalam 5 menit
13	Total kendaraan gol IV arah mendekat	<i>Int(5)</i>	Berisi total golongan IV arah mendekat atau arah ke Cirebon dalam 5 menit
14	Total kendaraan gol V arah mendekat	<i>Int(5)</i>	Berisi total golongan V arah mendekat atau arah ke Cirebon dalam 5 menit
15	Volume lalu lintas	<i>Int(5)</i>	Volume lalu lintas dengan satuan smp/jam
16	Kapasitas Jalan	<i>Int(5)</i>	Kapasitas jalan dengan satuan smp/jam

Dengan adanya nilai kapasitas jalan dan nilai volume lalu lintas dengan satuan smp/jam maka dari sini akan didapatkan nilai Volume kapasitas rasio (VCR). Untuk mendapatkan *field* kondisi lalu lintas harus mengetahui *Volume Kapasitas Rasio (VCR)* atau tingkat pelayanan jalan dengan cara data kapasitas jalan di bagi dengan data volume lalu lintas jika hasilnya lebih dari 1 maka dapat dipastikan kondisi lalu lintas sedang macet. Berikut struktur data lalu lintas setelah mencantumkan nilai VCR dan kondisi lalu lintas:

Tabel 4.4 Struktur Tabel Setelah Mencantumkan Nilai VCR dan Kondisi Lalu Lintas

No.	Field	Type	Keterangan
1	Id	<i>Int(8)</i>	Merupakan nomor urut
2	Lokasi	<i>Varchar (50)</i>	Berisi Lokasi perangkat terpasang
3	Waktu	<i>Timestamp</i>	Berisi waktu pengambilan data kendaraan yang lewat selama 5 menit
4	Jenis Jalan	<i>Varchar (20)</i>	Berisi Jenis jalan, misalkan dalam kota atau luar kota
5	Total kendaraan gol I arah menjauh	<i>Int(5)</i>	Berisi total golongan I arah menjauh atau arah ke Purwakarta dalam 5 menit
6	Total kendaraan gol II arah menjauh	<i>Int(5)</i>	Berisi total golongan II arah menjauh atau arah ke Purwakarta dalam 5 menit
7	Total kendaraan gol III arah menjauh	<i>Int(5)</i>	Berisi total golongan III arah menjauh atau arah ke Purwakarta dalam 5 menit
8	Total kendaraan gol IV arah menjauh	<i>Int(5)</i>	Berisi total golongan IV arah menjauh atau arah ke Purwakarta dalam 5 menit
9	Total kendaraan gol V arah menjauh	<i>Int(5)</i>	Berisi total golongan V arah menjauh / arah ke Purwakarta dalam 5 menit

No.	Field	Type	Keterangan
10	Total kendaraan gol I arah mendekat	Int(5)	Berisi total golongan I arah mendekat atau arah ke Cirebon dalam 5 menit
11	Total kendaraan gol II arah mendekat	Int(5)	Berisi total golongan II arah mendekat atau arah ke Cirebon dalam 5 menit
12	Total kendaraan gol III arah mendekat	Int(5)	Berisi total golongan III arah mendekat atau arah ke Cirebon dalam 5 menit
13	Total kendaraan gol IV arah mendekat	Int(5)	Berisi total golongan IV arah mendekat atau arah ke Cirebon dalam 5 menit
14	Total kendaraan gol V arah mendekat	Int(5)	Berisi total golongan V arah mendekat atau arah ke Cirebon dalam 5 menit
15	Volume lalu lintas	Int(5)	Volume lalu lintas dengan satuan smp/jam
16	Kapasitas Jalan	Int(5)	Kapasitas jalan dengan satuan smp/jam
17	VCR	Real(5)	Nilai Volume Kapasitas ratio
18	Y	Int(2)	Nilai -1 atau 1
19	Kondisi lalu lintas	Varchar (1)	Kondisi lalu lintas: macet atau tidak mace

Pada tabel 4.4 terdapat *field* waktu dengan *timestamp* nilai ini akan di *convert* untuk menghasilkan *field* baru yaitu hari, dengan adanya *field* hari ini akan bermanfaat untuk mencari hubungan data antara hari dengan tingkat kepadatan lalu lintas pada hari-hari tertentu.

Untuk meringkaskan tabel pada tabel 4.4 maka total kendaraan pada tiap golongan akan menjadi satu *field* karena dalam penelitian ini tidak memisahkan jumlah lajur mendekat ataupun lajur menjauh, jadi semuanya digabung menjadi satu. Sehingga struktur data lalu lintas setelah mendapatkan nilai hari dan di satukannya nilai jumlah kendaraan tiap golongan, berikut struktur tabelnya:

Tabel 4.5 Struktur Tabel Setelah Mencantumkan Nilai VCR dan Kondisi Lalu Lintas

No	Field	Type	Keterangan
1	Id	Int(8)	Merupakan nomor urut
2	Lokasi	Varchar (50)	Berisi Lokasi perangkat terpasang
3	Waktu	Timestamp	Berisi waktu pengambilan data kendaraan yang lewat selama 5 menit
4	Hari	Varchar (6)	Berisi hari: Senin, Selasa, Rabu, Kamis, Jumat, Sabtu dan Minggu

No	Field	Type	Keterangan
5	Jenis Jalan	<i>Varchar (20)</i>	Berisi Jenis jalan, misalkan dalam kota atau luar kota
6	Gol I	<i>Int(5)</i>	Berisi total golongan I penjumlahan dari arah menjauh dan arah mendekat dalam 5 menit
7	Gol II	<i>Int(5)</i>	Berisi total golongan II penjumlahan dari arah menjauh dan arah mendekat dalam 5 menit
8	Gol III	<i>Int(5)</i>	Berisi total golongan III penjumlahan dari arah menjauh dan arah mendekat dalam 5 menit
9	Gol IV	<i>Int(5)</i>	Berisi total golongan IV penjumlahan dari arah menjauh dan arah mendekat dalam 5 menit
10	Gol V	<i>Int(5)</i>	Berisi total golongan V penjumlahan dari arah menjauh dan arah mendekat dalam 5 menit
11	Volume lalu lintas	<i>Int(5)</i>	Volume lalu lintas dengan satuan smp/jam
12	Kapasitas Jalan	<i>Int(5)</i>	Kapasitas jalan dengan satuan smp/jam
13	VCR	<i>Real(5)</i>	Nilai Volume Kapasitas ratio
14	Y	<i>Int(2)</i>	Nilai -1 atau 1
15	Kondisi lalu lintas	<i>Varchar (1)</i>	Kondisi lalu lintas: macet atau tidak macet

4.4 Data Mining

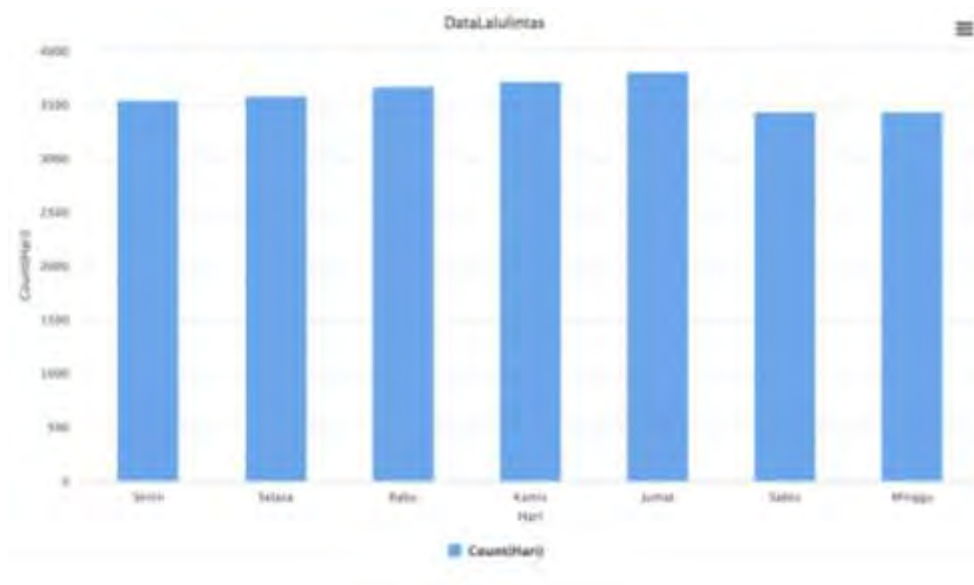
4.4.1 Analisis Deskriptif

Dalam analisis deskriptif membahas tentang merangkum sekumpulan data dalam bentuk yang mudah dibaca dan cepat memberikan informasi, yang disajikan dalam bentuk grafik dan nilai penyebarannya. Berikut gambaran data lalu lintas pada *software rapidminer*.

	Lokasi	Timestamp	Hari	jenis Jalan	GDL 1	GDL 2	GDL 3	GDL 4	GDL 5	Volume Jal.	Kapasitas Jal.	VCR	Kombi jal.
1	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	82	8	7	0	29	1645.200	2893	0.569	-1
2	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	49	1	8	0	4	446.800	2893	0.154	-1
3	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	83	24	12	0	24	3542.200	2893	0.535	-1
4	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	91	34	7	0	36	1881.200	2893	0.651	-1
5	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	84	24	6	0	32	1775.600	2893	0.613	-1
6	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	46	15	5	0	47	1862	2893	0.574	-1
7	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	37	15	1	0	33	1209.600	2893	0.418	-1
8	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	30	14	10	0	24	3248	2893	0.431	-1
9	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	81	28	12	0	35	1825.200	2893	0.631	-1
10	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	47	8	8	0	41	1302.400	2893	0.450	-1
11	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	87	62	4	0	38	1702.400	2893	0.589	-1
12	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	55	18	9	0	27	1310.400	2893	0.453	-1
13	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	55	92	12	0	36	2005.200	2893	0.693	-1
14	Jl. Raya By Pass Jk.	Mar 5, 2015...	Senin	Jalan Kota	72	22	10	0	23	1404	2893	0.485	-1

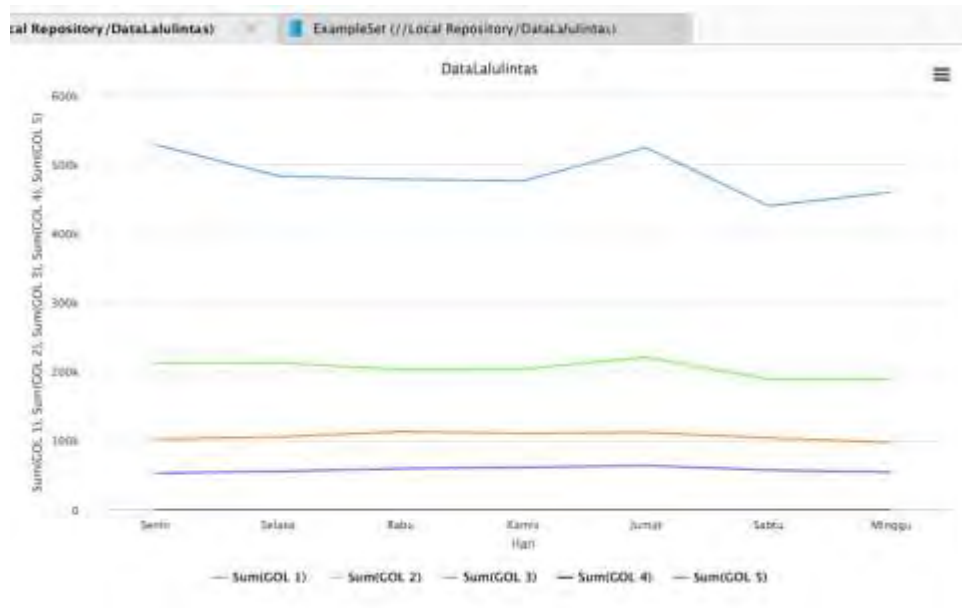
Gambar 4.3 View Data Lalu Lintas Pada Software Rapidminer

Terlihat pada gambar 4.3 tampilan data lalu lintas akan di olah dengan menggunakan *software Rapidminer*, penggunaan *software* ini untuk mempermudah dalam proses *data mining*.



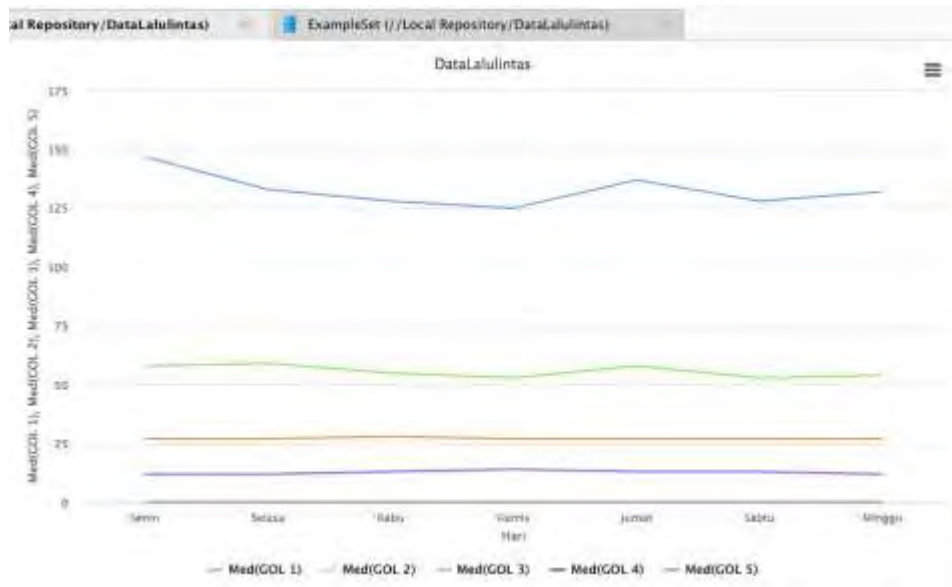
Gambar 4.4 Total Jumlah Kendaraan Per Hari Dari Data Lalu Lintas Yang Di Proses

Pada Gambar 4.4 merupakan grafik diagram batang gambaran total jumlah data lalu lintas yang diproses yang tersebar pada hari Senin, Selasa, Rabu, Kamis, Jumat, Sabtu dan Minggu.



Gambar 4.5 Grafik Jumlah Total kendaraan Per Hari Dari Setiap Golongan

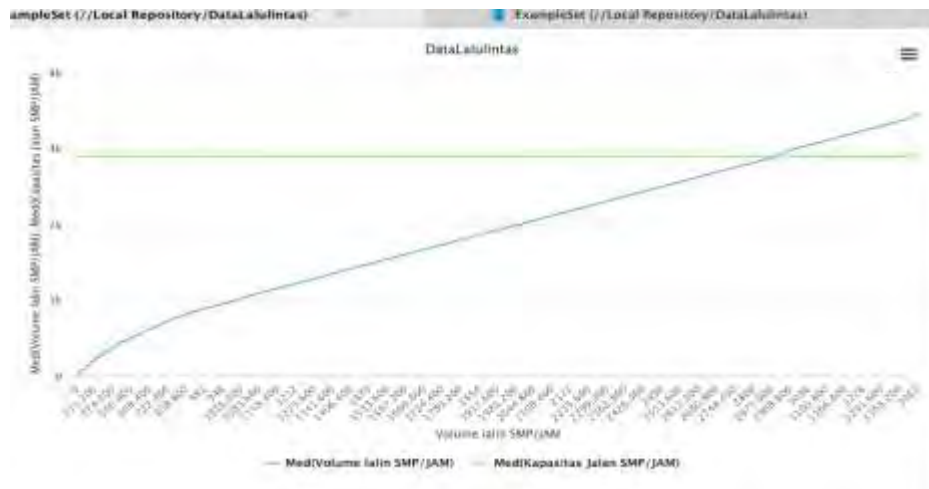
Dalam gambar 4.5 terlihat grafik perbedaan jumlah total kendaraan pada tujuh hari setiap golongan, pada grafik terlihat jumlah kendaraan golongan I atau sepeda motor mendominasi jumlahnya dibandingkan dengan jumlah kendaraan pada golongan II, golongan III, golongan IV dan golongan V.



Gambar 4.6 Jumlah rata-rata kendaraan lewat dalam 5 menit di hari tertentu

Terlihat pada gambar 4.6 jumlah rata-rata kendaraan yang lewat di hari tertentu, terlihat pada grafik kendaraan golongan I dalam 5 menit rata-rata berjumlah 130 kendaraan, selanjutnya golongan II dalam 5 menit rata-rata berjumlah 60 kendaraan,

golongan III dalam 5 menit rata-rata berjumlah 30 kendaraan, pada golongan V dan IV dalam 5 menit rata-rata berjumlah 12 kendaraan.

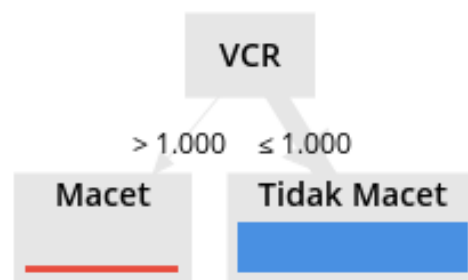


Gambar 4.7 Grafik perbandingan kapasitas jalan dengan volume lalu lintas

Gambar 4.7 menampilkan grafik perbandingan kapasitas jalan dengan volume lalu lintas berdasarkan data lalu lintas yang di proses, pada grafik menggambarnya adanya informasi bahwa pada data lalu lintas adanya kemacetan pada waktu tertentu.

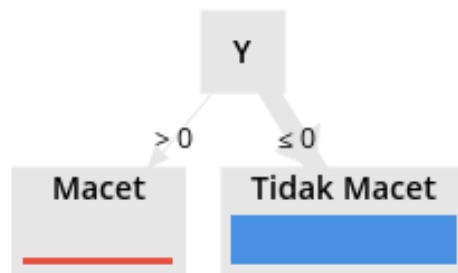
4.4.2 Random Forest

Dalam bab sebelumnya di bahas *random forest* merupakan pengembangan dari *decision tree* yaitu memiliki banyak *tree*. Berikut beberapa *tree* yang terbentuk oleh *random forest* :



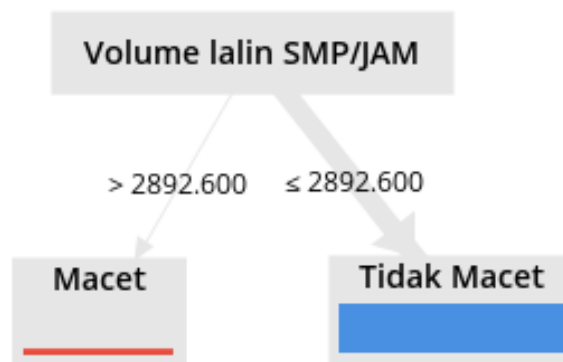
Gambar 4.8 Tree Nilai VCR Menentukan Macet

Gambar 4.8 menunjukkan bahwa nilai VCR (*Volume Capacity Rasio*) jika di atas angka 1(satu) akan mendapatkan informasi label macet sedangkan jika nilainya lebih kecil atau sama dengan 1 (satu) maka akan menunjukkan informasi label tidak macet.



Gambar 4.10 Tree Nilai Y Menentukan Macet

Pada Gambar 4.9 terlihat nilai Y jika di atas angka 0 (nol) akan mendapatkan informasi label macet sedangkan jika nilainya lebih kecil atau sama dengan 0 (nol) maka akan menunjukkan informasi tidak macet.



Gambar 4.10 Tree Nilai Volume Lalu Lintas

Gambar 4.10 terlihat nilai volume lalu lintas jika lebih dari atau sama dengan 2.892 smp/jam maka label informasi kondisi jalan akan menunjukkan bahwa jalan tersebut terjadi kemacetan jika nilai volume lalu lintas kurang dari 2.892 smp/jam maka informasi kondisi jalan akan menunjukkan label informasi bahwa jalan tersebut tidak macet.



Gambar 4.11 Tree Jam 05:30

Pada Gambar 4.11 *node* akar (*root*) adalah label jam ini menandakan jam merupakan atribut yang memiliki pengaruh terbesar. Jika jam menunjukkan kurang dari pukul 05:30 mempunyai peluang besar lokasi menunjukkan informasi label tidak macet, jika lebih dari pukul 05:30 dan nilai Y lebih dari 0 (nol) akan berpeluang menjadi informasi label macet sedangkan jika lebih dari pukul 05:30 dan nilai Y kurang dari 0 (nol) informasi label akan menunjukkan tidak macet.



Gambar 4.12 Tree Tingkat Kepadatan Lalu lintas

Pada subbab Analisa deskripsi di bahas kendaraan golongan I atau sepeda motor jumlahnya mendominasi dibandingkan dengan kendaraan golongan lain. *node*

akar (*root*) dalam gambar 4.12 adalah label Gol1 ini menandakan kendaraan sepeda motor memiliki pengaruh terbesar dalam membuat keluarnya label tidak macet. pada gambar 4.12 terlihat jika jumlah golongan I lebih kecil sama dengan 189,5 pada pukul 15:30 kurang dan jenis kendaraan golongan III nilainya kurang dari atau sama dengan 106,5 dan nilai VCR kurang dari sama dengan 1 (satu) maka dapat dipastikan informasi label lalu lintas akan menunjukkan tidak macet.

Dalam gambar 4.12 menunjukkan jika nilai kendaraan golongan I lebih kecil sama dengan 189,5 pada pukul kurang dari 15:30 di hari Senin, Selasa, Rabu, Kamis dan Sabtu berpeluang menunjukkan informasi label tidak macet, sedangkan pada hari Minggu lebih dari pukul 08:30 akan menunjukkan informasi label tidak macet dan pada hari Minggu berpeluang menunjukkan label informasi macet pada pukul kurang dari 18:30. Selain hari Minggu peluang keluarnya informasi label macet dapat terjadi di Jumat dan hari Senin, pada hari Senin peluang keluarnya informasi label macet terjadi jika jumlah golongan I bernilai 189,5 pada pukul kurang dari 15:30 dan jumlah kendaraan golongan III 106,5 lebih.

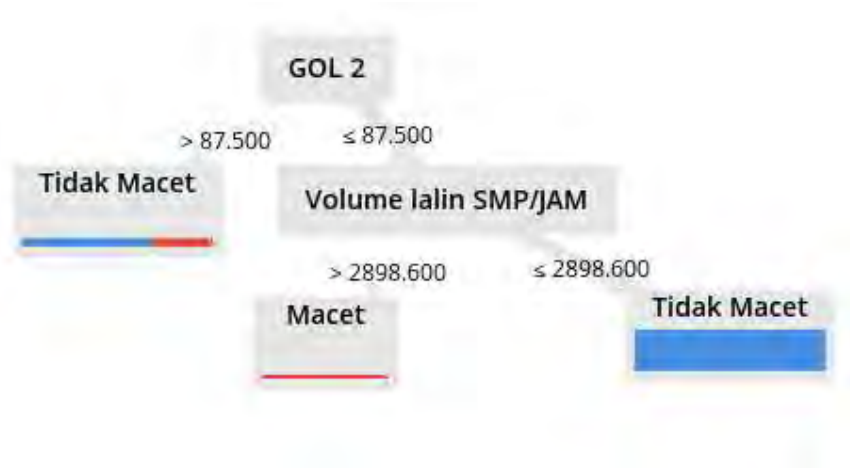
Gambar 4.12 juga menunjukkan peluang keluarnya informasi label tidak macet pada hari Selasa, Rabu, Kamis jika golongan III lebih dari 106,5 pada pukul 15:30 dan kendaraan golongan I bernilai kurang dari sama dengan 189,5.



Gambar 4.13 Tree Golongan I Dengan VCR

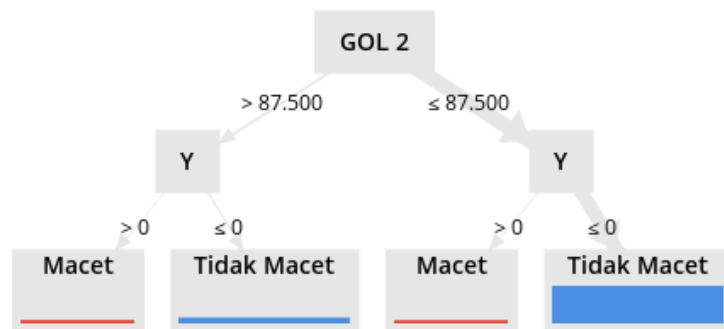
Seperti yang tergambar pada gambar 4.13 Informasi label tidak macet yang berpeluang besar keluar jika nilai VCR kurang dari sama dengan 1 (satu) dan nilai

golongan I kurang dari sama dengan 196,5. Informasi macet akan keluar saat nilai VCR lebih besar dari 1 (satu) saat golongan 1 bernilai golongan I kurang dari sama dengan 196,5.



Gambar 4.14 Tree Golongan 2 dengan Volume Lalu Lintas

Tree pada gambar 4.14 menunjukkan jika kendaraan golongan II juga mempunyai peranan penting dalam mempengaruhi macet atau tidaknya lokasi jalan. Jika nilai kendaraan golongan II kurang dari sama dengan 87,5 dan nilai volume lalu lintas kurang dari atau sama dengan 2898,6 maka informasi label lalu lintas tidak macet. Pada tree ini label Gol 2 menjadi *node* akar menunjukkan bahwa kendaraan dengan golongan 2 mempengaruhi keluarnya label macet dan tidak macet.



Gambar 4.15 Tree Golongan 2 Dengan Y

Pada gambar 4.15 kendaraan golongan II akan mempengaruhi keluarnya informasi tidak macet jika nilai golongan kelas II kurang dari sama dengan 87,5 dan nilai Y lebih kecil sama dengan 0 (nol).



Gambar 4.16 Tree Jam Dengan Y

Gambar 4.16 menunjukkan jam berpengaruh terhadap keluarnya informasi label macet atau tidaknya lokasi jalan. Saat jam menunjukkan pukul kurang dari 05:30 maka label informasi akan keluar tidak macet, ini menunjukkan bahwa lokasi saat subuh jam 05:30 lokasi jalan *By Pass* Jomin tidak terjadi kemacetan pada hari Senin, Selasa, Rabu, Kamis, Jumat Sabtu dan Minggu.

Pada gambar 4.16 peluang keluarnya informasi label macet terjadi saat pukul 05:30 lebih dengan nilai golongan V lebih kecil atau sama dengan 93 pada tanggal 5 Maret 2018 di bawah jam 10:00 dan ini merupakan hari Senin.

4.4.2 Performance Random Forest

Berdasar hasil analisis data menggunakan *rapidminer* berikut *performance Random Forest* untuk klasifikasi data lalu lintas:

Performances

Criterion	Value	Standard Deviation
Accuracy	95.6%	± 0.3%
Classification Error	4.4%	± 0.3%
AUC	99.9%	± 0.0%
Precision	92.8%	± 0.5%
Recall	100.0%	± 0.0%
F Measure	96.3%	± 0.3%
Sensitivity	100.0%	± 0.0%
Specificity	89.9%	± 0.6%

Gambar 4.17 Performance Random Forest

Gambar 4.17 menunjukkan tingkat akurasi *Random Forest* pada pengolahan untuk klasifikasi data lalu lintas ini mempunyai akurasi sebesar 95,6%, tingkat *error* didapatkan sebesar 4,4% dan nilai *recall* 100%.

Komparasi *data mining* klasifikasi dengan menggunakan *tools automodel* dari *rapidminer*. Berikut hasilnya:



Gambar 4.18 Komparasi Random Forest Dengan

Pada gambar 4.18 pemrosesan data lalu lintas *Random Forest* memang mempunyai nilai *accuracy* lebih tinggi di banding *data mining* klasifikasi yang lain seperti *Naive Bayes* dan *Decision Tree*. *Accuracy Naive bayes* bernilai 95.1%, *Decision Tree* 86,2% dan *Random Forest* di 95,6%

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan mengenai klasifikasi tingkat kepadatan lalu lintas oleh pengguna kendaraan bermotor dengan menggunakan *Random Forest* pada jalan raya *By Pass* Jomin Cikampek, maka dapat diambil beberapa kesimpulan yaitu:

1. Setelah menerapkan klasifikasi dengan *Random Forest* hasil klasifikasi menunjukan beberapa kondisi macet dan tidak macet. Kondisi macet terjadi pada hari Senin disebabkan oleh golongan I ≥ 190 kendaraan dan golongan III ≥ 107 kendaraan terjadi antara 05:30 sampai dengan 15:30. Sedangkan kondisi macet yang terjadi pada hari Jumat disebabkan oleh golongan I ≥ 190 kendaraan dan golongan III ≥ 107 kendaraan terjadi saat $\geq 15:30$ serta volume lalu lintas bernilai > 2950 smp/jam. Selain hari Senin dan Jumat, kondisi lalu lintas macet juga terjadi di hari Minggu disebabkan oleh golongan I ≥ 190 kendaraan saat 15:30 sampai 18:30. Sedangkan kondisi lalu lintas cukup lancar terjadi pada hari Selasa, Rabu, Kamis dan Sabtu saat jam 00:00 sampai 15:30.
2. Tingkat akurasi *Random Forest* pada data lalu lintas ini mempunyai akurasi sebesar 95,6% tingkat *error* didapatkan sebesar 4,4% dan nilai *recall* 100%.
3. Ruas jalan yang di teliti terjadi kemacetan pada hari dan jam tertentu masukan untuk para *stakholder* untuk mengatur lalu lintas sebelum terjadi kemacetan atau memperluas jalan tersebut menjadi empat lajur tak terbagi (4/2 UD)

5.2 Saran

Dari hasil kesimpulan dapat diberikan saran-saran bagi perbaikan klasifikasi tingkat kepadatan lalu lintas oleh pengguna kendaraan bermotor dengan menggunakan *random forest*. Penelitian ini masih jauh dari sempurna, berikut saran bagi yang akan mengembangkan penelitian ini:

1. Melakukan penelitian tentang *data mining* untuk memprediksi kemacetan

2. Menganalisis derajat kejenuhan dan biaya kemacetan
3. Menganalisis biaya kemacetan yang ditanggung oleh pengguna jalan mobil pribadi dan masyarakat secara luas sebagai dampak dari kemacetan yang terjadi
4. Perlu dilakukan penelitian lanjutan untuk mengukur derajat kejenuhan pada jalan-jalan lainnya, dalam rangka perbaikan dan peningkatan pelayanan transportasi darat.

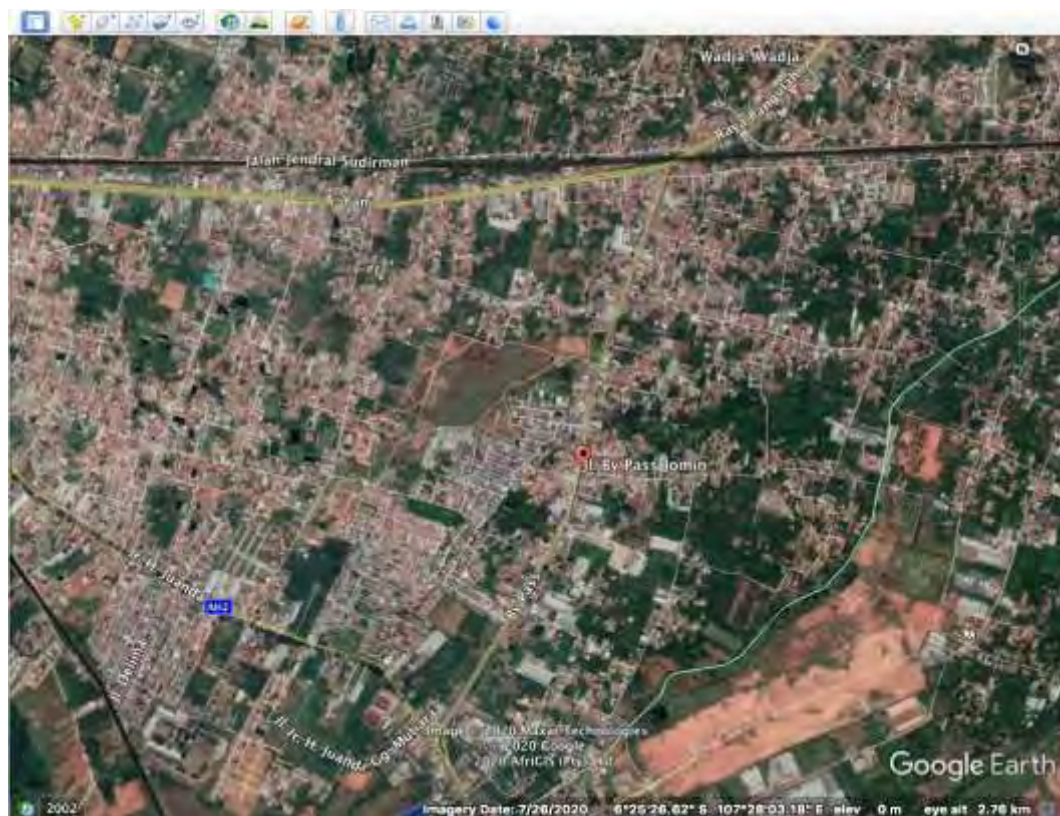
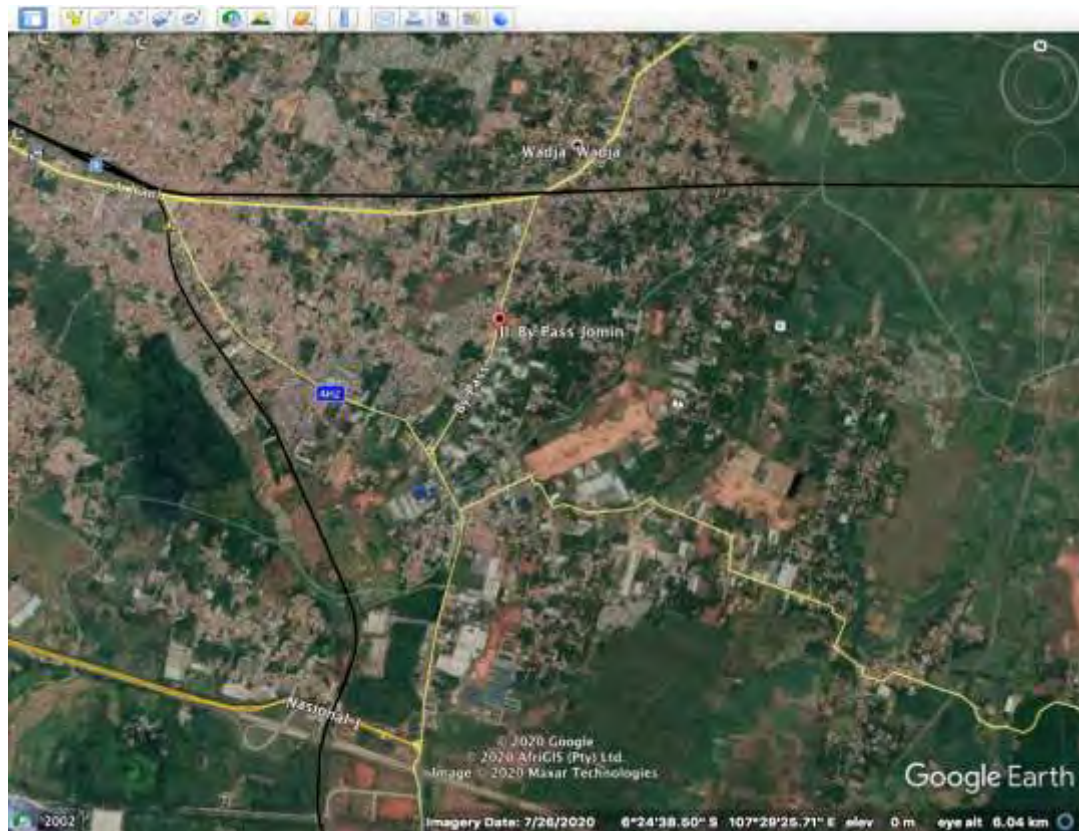
DAFTAR PUSTAKA

- Adnyana. Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus: Stikom Bali), Stikom Bali. Bali. 2016
- Antara, <https://www.beritasatu.com/nasional/236320-arus-simpang-jominjalur-pantura-karawang-padat>. Berita satu.com.Karawang.2014(diakses: 23 Juni 2020)
- Berry, Linoff. Data Mining Techniques Third Edition. Wiley Publishing Inc. 2011
- Biau, Gerard. Analysis of a Random Forests Model, Universitas Piere. France. 2012
- Breiman, Leo. Random Forests, University of California, Berkeley. 2001
- Breiman, Leo. Classification and Regresstion Trees. Chapmen&Hall. 2017
- Castro, Ester. Evaluasi Kinerja Ruas Jalan Audian Dili Timor Leste. Universitas Atma Jaya Yogyakarta. Yogyakarta. 2014
- Denisko, Danielle, and Hoffman. Classification and Interaction In Random Forests, PNAS. 18000256115, Canada. 2018
- Departemen Pekerjaan Umum, Manual Kapasitas Jalan Indonesia, Jakarta, Ditjen Bina Marga, 1997
- Han, Kamber, And Pei. Data Mining: Concept and Techniques, Third Edition. Waltham: Morgan Kaufmann Publishers. 2012
- Hanafiah, Disi Mochamad, Ariephin, Sailendra. Teknologi Pemrosesan Gambar Video Untuk Perhitungan Arus Sepeda Motor. Puslitbang Jalan dan Jembatan, Bandung. 2014
- Hanafiah, Disi Mochamad. Akurasi Alat Penghitung Lalu Lintas Plato 2.1 Berbasis Pengolahan Citra - Background Substraction. Puslitbang Jalan dan Jembatan, Bandung. 2018
- Mambang, Mambang dan Byna. Analisis Perbandingan Algoritma C.45, Random Forest Dengan Chaid Decision Tree Untuk Klasifikasi Tingkat Kecemasan Ibu Hamil, Universitas Amikom, Yogyakarta. 2017
- Muslikh, Ahmad, Santoso, Marjuni. Klasifikasi Data Time Series Arus Lalu Lintas Jangka Pendek Menggunakan Algoritma Adaboost Dengan Random Forest. Jurnal Teknologi Informasi Volume 14 Nomor 1 ISSN 1907-3380, Semarang. 2019
- Nidhomuddin and B. W. Otok. Random Forest Dan Multivariate Adaptive Regression Spline (Mars) Binary Response Untuk Klasifikasi Penderita Hiv / Aids Di Surabaya. Statistika Vol. 1 No. 3. Surabaya. 2015
- Saleh, Alfa. Penerapan Data Mining Dengan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Kelulusan Mahasiswa Dalam Mengikuti English Proficiency Test. Teknik Informatika Universitas Potensi Utama. 2015
- Saputra, Rizal Amegia. Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi. Seminar Nasional : Program Studi Manajemen Informatika. Sukabumi. 2014
- Shandi, Yusup Jauhari. Analisis Sistem Perencanaan Transportasi Umum Dengan

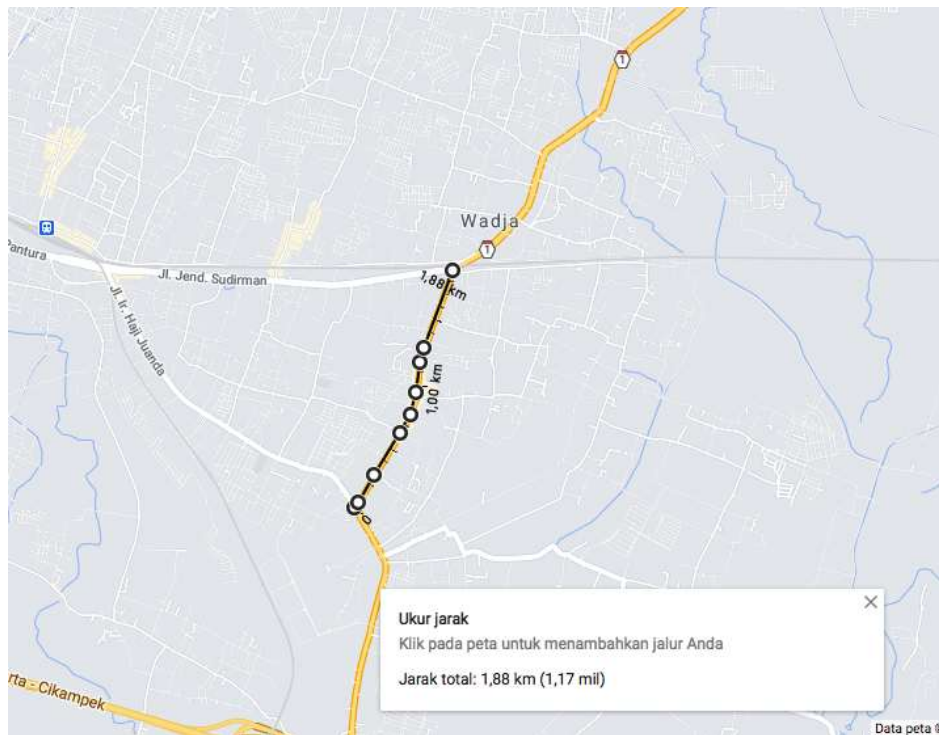
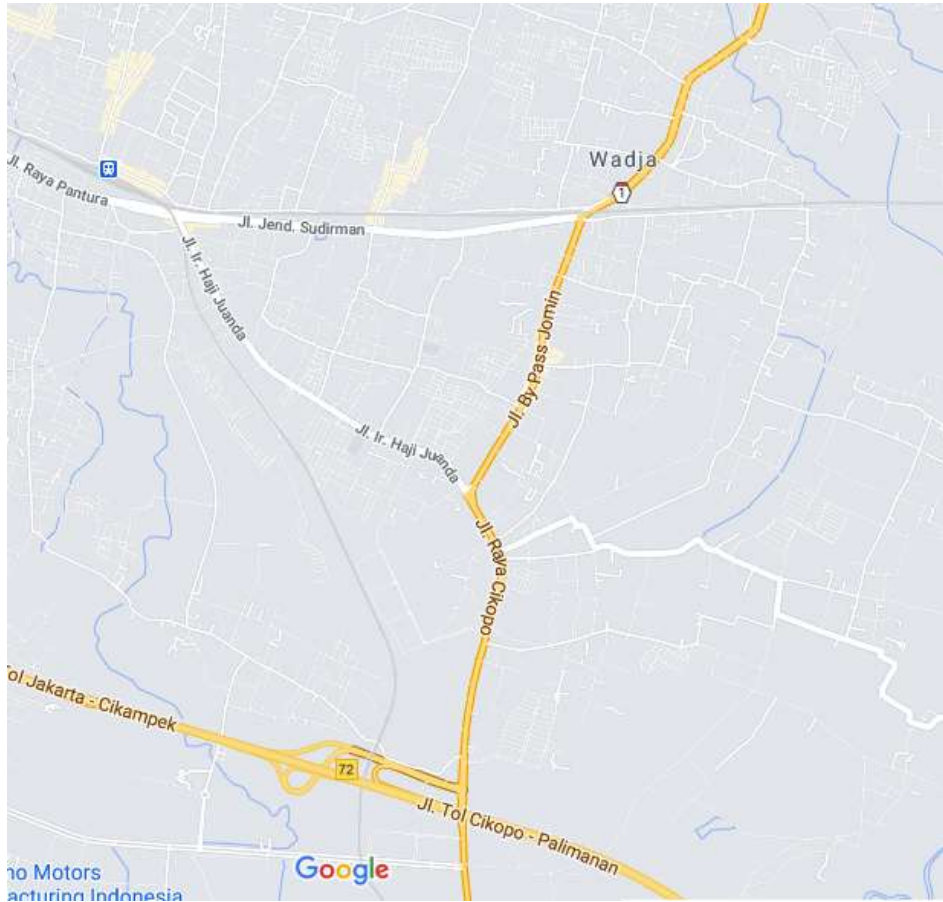
- Menggunakan Model Volume Capacity Rasio. STMIK LIKMI. 2018
- Shella, pradega. Sistem Pendukung Keputusan Dengan Menggunakan Decision Tree Dalam Pemberian Beasiswa Di Sekolah Menengah Pertama. UNNES. 2015
- Sudarta, Dwi. Pengaruh Sistem kegiatan Serta Keberadaan Terminal Bayangan Terhadap Kinerja Jalan Raden Intan Kota Bandar Lampung. PWK UNIKOM. 2013
- Sugiyono. Metode Penelitian Kuantitatif, Kualitatif dan R&D, Alfabeta, Bandung. 2011
- Swastina, L. Penerapan Algoritma C4.5 Untuk Penentuan Jurusan Mahasiswa. Gema Aktualita, 2(1), 93-98, Surabaya. 2013
- Touw, Wouter G, and Bayjanov. Data Mining In The Life Sciences With Random Forest: A Walk In The Park Or Lost In The Jungle?, Briefings In Bioinformatics, Inggris. 2012
- Verikas, A, Gelzinis, and Bacauskiene. Mining data with random forests: A survey and results of new tests. Elsevier, Sweden. 2011
- Wezel, Potharst. Improved Customer Choice Predictions using Ensemble Methods. European Journal of Operational Research. 2007
- Widiastuti, Julia. Klasifikasi Pembiayaan Warung Mikro Menggunakan Metode *Random Forest* dengan Teknik Sampling kelas Imbalanced, Statistika UII. Yogyakarta. 2018
- Witten IH, Frank E. Data Mining: Practical Machine Learning Tools and Techniques 2nd edition. Morgan Kaufmann Publishers, San Francisco. 2016

LAMPIRAN

LAMPIRAN A

Peta Lokasi Jalan Raya *By Pass* Jomin

Peta Lokasi Jalan Raya By Pass Jomin



Peta Lokasi Jalan Raya *By Pass* Jomin

Simpang tiga ke *by pass* jomin dari exit Tol Cikampek



Simpang tiga setelah *by pass* jomin dari exit Tol Cikampek



Peta Lokasi Jalan Raya *By Pass* Jomin

Simpang tiga *Exit* Tol Cikampek



Persimpangan dari Karawang menuju Jalan *By Pass* Simpang Jomin



