

**KLASIFIKASI KEJADIAN KECELAKAAN LALU LINTAS
BERDASARKAN LUCA KORBAN DENGAN METODE SUPPORT
VECTOR MACHINE POLYNOMIAL SMOTE**

(Studi Kasus: Kecelakaan Lalu Lintas di Sleman, Yogyakarta Tahun 2016-2017)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Jurusan
Statistika



Disusun Oleh:

Indra Syaputra

13 611 196

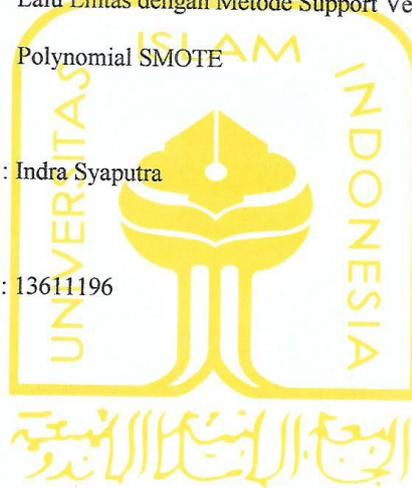
**JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2018**

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : Klasifikasi Tingkat Keparahan Luka Korban Kecelakaan
Lalu Lintas dengan Metode Support Vector Machine
Polynomial SMOTE

Nama Mahasiswa : Indra Syaputra

Nomor Mahasiswa : 13611196



**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 22 Februari 2018

Pembimbing I

A handwritten signature in black ink, appearing to be 'Muhajir', is written over a horizontal line.

(Muhammad Muhajir S.Si.,M.Sc.)

HALAMAN PENGESAHAN
TUGAS AKHIR
KLASIFIKASI KEJADIAN KECELAKAAN LALU LINTAS
BERDASARKAN LUKA KORBAN DENGAN METODE SUPPORT
VECTOR MACHINE POLYNOMIAL SMOTE
(Studi Kasus: Kecelakaan Lalu Lintas di Sleman, Yogyakarta
Tahun 2016-2017)

Nama : Indra Syaputra

Nim : 13611196

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL 19 Maret 2018

Nama Penguji

1. Ayundyah Kesumawati, M.Si
2. Ir. Ali Parkhan, M.T
3. Muhammad Muhajir S.Si, M.Sc

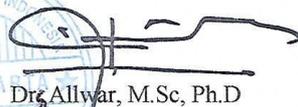
Tanda Tangan



Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam




Dr. Allwar, M.Sc, Ph.D

iii

KATA PENGANTAR

Bismillaahirrahmaanirrahiim.

Assalamu'alaikum Wr. Wb.

Puji dan syukur kehadirat Allah SWT yang telah melimpahkan rahmat, taufiq dan hidayah-Nya, sehingga penulis dapat menyusun dan menyelesaikan penulisan tugas akhir ini. Shalawat serta salam kepada Nabi Muhammad SAW beserta keluarga dan para pengikutnya sampai akhir zaman.

Tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana jurusan statistika di Universitas Islam Indonesia. Tugas akhir ini dilaksanakan pada bulan Oktober 2017 - Februari 2018 di Yogyakarta. Tugas akhir yang berjudul **“KLASIFIKASI KEJADIAN KECELAKAAN LALU LINTAS BERDASARKAN LUKA KORBAN DENGAN METODE SUPPORT VECTOR MACHINE POLYNOMIAL SMOTE”** ini selain disusun guna memenuhi persyaratan untuk menyelesaikan studi jenjang strata satu di Jurusan Statistika Universitas Islam Indonesia, juga untuk memberikan wawasan dan mengenalkan aplikasi metode statistik dalam kegiatan mitigasi bencana kepada masyarakat luas.

Selama mengerjakan tugas akhir ini, penulis telah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Untuk itu pada kesempatan ini penulis bermaksud menyampaikan ucapan terima kasih kepada :

1. Kedua orang tua tercinta, Bapak Ir. Indermawan, HB dan Ibu Ir. Baiq Rahma Iryanti yang telah mencurahkan kasih sayangnya, memberikan doa, batuan moril maupun materil, serta saudara terkasih Kakak tercinta Indria Syaputri dan Mas Eka yang selalu selalu memberikan doa, motivasi dan nasehat sehingga dapat memberikan semangat untuk pantang menyerah.
2. Bapak Dr. RB. Fajriya Hakim, M.Si. selaku Ketua Jurusan Statistika FMIPA UII beserta jajarannya.
3. Bapak Muhammad Muhajir S.Si., M.S.i selaku dosen pembimbing yang sangat sabar dan berjasa membimbing dalam penyelesaian tugas akhir ini.

4. Seluruh Dosen Pengajar dan Staff Prodi Statistika yang berkontribusi dalam membantu penulis selama masa perkuliahan.
5. Nur Azirah yang telah bersedia menjadi *support system* bagi penulis serta sabar dalam memberikan semangat untuk penyelesaian penulisan tugas akhir ini.
6. Teman-teman Statistika angkatan 2013 dan teman-teman statistika kelas D yang selalu memberikan motivasi untuk selalu berjuang.
7. Teman-teman LASER-C periode 2015-2016 yang sudah membagi ilmu dan pengalaman untuk penulis.
8. KKN Unit 107 Keburusan, Purworejo, Adi, Oky, Firzha, Fadhil, Tommy, Siti dan Dwi yang telah memberikan doa serta hiburan kepada penulis.
9. Teman seperjuangan sekaligus teman seataap selama kuliah, Rikky dan Isna. Terima kasih atas segala dukungan.
10. Bapak dan Ibu Anto yang selama ini sudah menjadi orang tua substitusi selama menjalani masa perkuliahan diperantauan.
11. Bapak Nurdiatno dan segenap instansi terkait yang sudah membantu dan meluangkan waktunya demi terselesaikannya tugas akhir ini.
12. Semua pihak yang telah mendukung dan ikut berpartisipasi penulis ucapkan banyak terima kasih.

Semoga dukungan dan bantuan dari semua pihak senantiasa mendapat balasan yang dari Allah SWT. Penulis menyadari bahwa tugas akhir ini masih memiliki kekurangan di dalamnya. Hal tersebut karena keterbatasan ilmu dan pengetahuan yang dimiliki penulis semata. Penulis menerima kritik dan saran yang membangun demi perbaikan tugas akhir ini. Penulis berharap agar penelitian ini dapat bermanfaat dan memberikan khasanah pengetahuan bagi penulis, pembaca, maupun penelitian di masa depan.

Wassalamu'alaikum Wr. Wb.

Yogyakarta, 22 Februari 2018

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
PERNYATAAN	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
BAB III LANDASAN TEORI	9
3.1. Kecelakaan Lalu Lintas	9
3.1.1. Pengertian Kecelakaan Lalu Lintas	9
3.1.2. Karakteristik Kecelakaan Lalu Lintas	9
3.1.3. Faktor-faktor Penyebab Kecelakaan Lalu Lintas	10
3.2. <i>Data Mining</i>	11
3.3. Vektor	13
3.4. Support Vector Machine (SVM)	15
3.4.1. SVM Pada Data Terpisah Secara <i>Linear</i>	16
3.4.2. SVM Pada Data Tidak Terpisah Secara Linear	19
3.4.3. <i>Kernel Trick</i> dan <i>Non-Linear Classification</i> pada SVM..	19

3.5 Accuracy dan Precision.....	22
3.6. Imbalance Dataset.....	23
3.6.1. SMOTE (<i>Synthetic Minority Over-sampling Technique</i>)	29
3.7. Confussion Matrix	26
BAB IV METODOLOGI PENELITIAN	28
4.1. Populasi dan Sampel Penelitian	28
4.2. Sumber Data	28
4.3. Variabel Penelitian dan Definisi Operasional Variabel	28
4.4. Metode Penelitian	30
4.5. Langkah Penelitian	30
BAB V HASIL DAN PEMBAHASAN	32
5.1. Analisis Deskriptif.....	32
5.2. Analisis Data dengan Metode SVM.....	39
5.2.1. Klasifikasi menggunakan <i>SVM polynomial Kernel</i>	39
5.2.2. Klasifikasi menggunakan <i>SVM polynomial Kernel</i> dengan metode <i>SMOTE</i>	41
5.3. Validasi Metode.....	45
BAB VI KESIMPULAN DAN SARAN	46
6.1. Kesimpulan	46
6.2. Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN	50

DAFTAR TABEL

Nomor	Judul	Halaman
1.1	Kecelakaan Lalu Lintas Sepeda Motor menurut Kabupaten/Kota di D.I Yogyakarta Tahun 2004-2015	1
3.1	<i>Confusion Matrix</i>	26
5.1	Crosstab Jumlah Tingkat Keparahan Luka Menurut Waktu Kejadian	36
5.2	Crosstab Jumlah Tingkat Keparahan Luka Menurut Usia Korban	36
5.3	Crosstab Jumlah Tingkat Keparahan Luka Menurut Kepemilikan SIM	37
5.4	Crosstab Jumlah Tingkat Keparahan Luka Menurut Jenis Kelamin	37
5.5	Crosstab Jumlah Tingkat Keparahan Luka Menurut Jenis Kecelakaan	38
5.6	Crosstab Jumlah Tingkat Keparahan Luka Menurut Kendaraan Terlibat	39
5.7	Matrik Konfusi	40
5.8	Hasil Akurasi Klasifikasi SVM Kernel <i>Polynomial</i> pada Data <i>Training</i>	42
5.9	Hasil Akurasi Klasifikasi SVM Kernel <i>Polynomial</i> pada Data <i>Testing</i>	42
5.10	Matrik Konfusi SVM <i>Polynomial</i> Dengan SMOTE pada Data Training	43
5.11	Matrik konfusi SVM <i>Polynomial</i> dengan SMOTE pada data testing	44
5.12	Perbandingan Tingkat Akurasi	45
5.13	Perbandingan Tingkat Presisi	45

DAFTAR GAMBAR

Nomor	Judul	Halaman
3.1	Langkah-langkah untuk melakukan <i>Data Mining</i>	11
3.2	Vektor	13
3.3	Cara Segitiga	14
3.4	Aturan Jajar Genjang	14
3.5	Perkalian skalar dua vektor jika membentuk sudut	14
3.6	Jumlah Koperasi Tidak Aktif	16
3.7	Fungsi Φ memetakan data ke ruang vector yang berdimensi lebih tinggi, sehingga kedua kelas dapat dipisahkan secara <i>linear</i> oleh sebuah <i>hyperplane</i>	20
3.8	Akurasi Rendah, Presisi Rendah	22
3.9	Akurasi Rendah, Presisi Tinggi	23
3.10	Akurasi Tinggi, Presisi Rendah	23
3.11	Akurasi Tinggi, Presisi Tinggi	23
3.12	Ilustrasi SMOTE	26
4.1	Alur Penelitian	31
5.1	Jumlah Kasus Korban Kecelakaan Tahun 2016-2017 Menurut Tingkat Keparahan Luka	32
5.2	Waktu Kejadian	33
5.3	Usia Korban	33
5.4	Kepemilikan Surat Izin Mengendarai (SIM)	34
5.5	Jenis Kelamin	34
5.6	Jenis Kecelakaan	35
5.7	Kendaraan Terlibat	35
5.8	Ilustrasi perubahan jumlah data setelah oversampling menggunakan SMOTE	41

DAFTAR LAMPIRAN

- Lampiran 1 Data Kecelakaan Lalu Lintas Sleman, Yogyakarta Tahun 2016-2017
- Lampiran 2 *Output* Klasifikasi SVM Polynomial menggunakan Aplikasi WEKA

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 22 Februari 2018



Penulis

**KLASIFIKASI TINGKAT KEPARAHAN LUKA KORBAN
KECELAKAAN LALU LINTAS DENGAN METODE SUPPORT VECTOR
MACHINE
POLYNOMIAL SMOTE**

(Studi Kasus: Kecelakaan Lalu Lintas di Sleman, Yogyakarta Tahun 2016-2017)

Indra Syaputra
Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

INTISARI

Kecelakaan Lalu Lintas adalah suatu peristiwa di Jalan yang tidak diduga dan tidak disengaja melibatkan Kendaraan dengan atau tanpa Pengguna Jalan lain yang mengakibatkan korban manusia dan atau kerugian harta benda. Hampir 27.000 orang meninggal dunia akibat kecelakaan lalu lintas sepanjang tahun 2015 di Indonesia. Provinsi Daerah Istimewa Yogyakarta yang menduduki urutan ke-7 di Indonesia (BPS, 2015). Di DIY, Kabupaten Sleman merupakan Kabupaten dengan kasus kecelakaan yang memiliki kerugian terbesar dengan korban yang mengalami kecelakaan tersebut mengalami luka ringan, luka berat hingga kematian terbesar dari Kabupaten/Kota lainnya di Provinsi Yogyakarta. Jumlah korban dalam kasus kecelakaan ini tentunya sangat merugikan berbagai pihak dan apabila tidak diketahui penyebab utamanya bisa berdampak fatal. Oleh karena itu, peneliti mengklasifikasi tingkat keparahan luka korban kecelakaan lalu lintas untuk mengetahui segmen kejadian kecelakaan lalu lintas yang menyebabkan korban luka berat atau meninggal dunia.. Metode yang digunakan adalah metode SVM (Support Vector Machine) Polynomial Kernel SMOTE.. Data yang digunakan berasal dari Polres Sleman tahun 2016-2017 . Berdasarkan hasil klasifikasi SVM Polynomial SMOTE menggunakan data testing didapatkan 795 data dan ada sebanyak 654 data yang berhasil diklasifikasi dengan benar sedangkan sebanyak 141 data tidak diklasifikasi dengan benar. Berdasarkan perhitungan bahwa persentase ketepatan klasifikasi atau nilai keakuratan dalam menganalisis data keparahan luka korban kecelakaan lalu lintas Kabupaten Sleman adalah sebesar 82.264 % dengan tingkat kesalahan klasifikasi sebesar 17.735 % .

Kata kunci: *Kecelakaan Lalu Lintas, , Klasifikasi, Support Vector Machine (SVM), Polynomial SMOTE.*

***THE CLASSIFICATION OF THE SEVERITY OF THE INJURED VICTIMS
OF TRAFFIC ACCIDENTS USING POLYNOMIAL SMOTE SUPPORT
VECTOR MACHINE METHOD***

(Case Study: Sleman Yogyakarta's Traffic Accidents in 2016-2017)

Indra Syaputra

Department of Statistics, Faculty of Mathematics and Science

Islamic University of Indonesia

ABSTRACT

Traffic Accidents are unexpected and accidental incidents on the Road involving Vehicles with or without other Road User resulting in human casualties and / or loss of property. Nearly 27,000 people died from traffic accidents throughout 2015 in Indonesia. Province of Special Region of Yogyakarta which is ranked 7th in Indonesia (BPS, 2015). In DIY, Sleman regency is a district with an accident case that has the greatest losses with the casualties who suffered minor injuries, serious injuries to the greatest deaths from other districts in Yogyakarta Province. the number of casualties in the case of this accident would be very detrimental to various parties and if not known the main cause can be fatal impact. Therefore, the researcher classifies the severity of the injured victims of traffic accidents to provide information on the parties involved in order to reduce the number of accidents and traffic accident victims in Sleman District in the next year. The method used is SVM (Support Vector Machine) Polynomial Kernel SMOTE method .. The data used comes from Polres Sleman in 2016-2017. Based on the classification of SVM Polynomial SMOTE using data testing obtained 795 data and there are as many as 654 data successfully classified correctly while as many as 141 data are not classified correctly. Based on the calculation that the percentage of accuracy of classification or accuracy value in analyzing data on the severity of injured victims of traffic accident in Sleman Regency is 82.264% with classification error level is 17.735% .

Keyword: *Traffic Accident, Clasification, Support Vector Machine (SVM), Polynomial SMOTE.*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

“Kecelakaan Lalu Lintas adalah suatu peristiwa di Jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan korban manusia dan atau kerugian harta benda” (Undang-Undang Republik Indonesia Nomor 22, 2009). Berbagai akibat kecelakaan meliputi korban yang meninggal dunia, luka-luka berat dan ringan, serta kerugian materiil yang tidak kecil jumlahnya (Adisasmita, 2011). “Secara umum kecelakaan lalu lintas yang terjadi disebabkan oleh beberapa faktor, seperti kelalaian manusia, kondisi jalan, kelayakan kendaraan dan belum optimalnya penegakan hukum lalu lintas”.

Ada sekitar 752 pejalan kaki dan 786 pengguna kendaraan bermotor tewas di dunia per harinya (WHO, 2015). Ada beberapa negara yang masih memiliki tingkat kematian akibat kecelakaan lalu lintas dengan tingkat kematian yang tinggi seperti yang dilansir Global Status Report on Road Safety pada tahun 2015 , diantaranya adalah negara Tiongkok, India, Nigeria, Brazil dan Indonesia. “Kecelakaan lalu lintas di Indonesia oleh WHO dinilai menjadi pembunuh terbesar ketiga, di bawah penyakit jantung koroner dan Tuberculosis (TBC)”. “Berdasarkan data Korps Lalu Lintas (Korlantas) Polri, angka kecelakaan lalu lintas di Indonesia sepanjang 2015 sedikit turun dari 2014. Meski begitu jumlahnya masih tergolong tinggi. Kepala Korlantas Polri Irjen Pol Condro Kirono mengatakan hampir 27.000 orang meninggal dunia akibat kecelakaan lalu lintas. Dari jumlah tersebut, mayoritas korban adalah pengendara sepeda motor dengan usia produktif”.

Di Indonesia ada beberapa Provinsi dengan catatan angka kecelakaan yang tinggi, salah satunya Provinsi Daerah Istimewa Yogyakarta yang menduduki urutan ke-7 setelah Jawa Timur, Jawa Tengah, Jawa Barat, DKI Jakarta, Sumatera Utara dan Sulawesi Selatan (BPS, 2015). Sebagai kota budaya, DIY dikenal juga sebagai kota pelajar. Jumlah penduduk di DIY semakin bertambah begitu pula dengan keadaan lalu lintasnya yang mayoritas pengguna sepeda motor.

Tabel 1.1 Kecelakaan Lalu Lintas menurut Kabupaten/Kota di D.I Yogyakarta
(Number of Accident by Regency/ City in D.I Yogyakarta) Tahun 2015

Kabupaten/Kota	Jumlah Kecelakaan	Meninggal Dunia	Luka Berat	Luka Ringan
1. Kulonprogo	433	43	15	846
2. Bantul	1562	122	1	1804
3. Gunungkidul	407	28	3	564
4. Sleman	1260	160	30	1850
5. Yogyakarta	651	45	0	899

Sumber : Kantor Ditlantas Polda D.I. Yogyakarta

Menurut data BPS Provinsi DIY di atas, Kabupaten dengan jumlah kecelakaan lalu lintas terbanyak adalah Kabupaten Bantul dengan 1.562 jumlah kecelakaan, kemudian diikuti oleh Kabupaten Sleman dengan angka 1.260 kecelakaan, Kota Yogyakarta dengan jumlah 651 kasus kecelakaan, Kabupaten Kulonprogo dengan jumlah 433 kasus kecelakaan dan Kabupaten Gunungkidul dengan jumlah 407 kasus kecelakaan.

Dari seluruh data jumlah kecelakaan pada seluruh Kabupaten di Provinsi DIY, Kabupaten Sleman merupakan Kabupaten dengan kasus kecelakaan yang memiliki kerugian terbesar dengan korban yang mengalami kecelakaan tersebut mengalami luka ringan, luka berat hingga kematian terbesar dari Kabupaten/Kota lainnya di Provinsi Yogyakarta. Tercatat 160 korban jiwa, 30 korban luka berat dan

1.850 korban luka ringan yang disebabkan oleh kasus kecelakaan di Kabupaten Sleman, Yogyakarta. Pada Tahun 2016, jumlah kecelakaan di Kabupaten Sleman meningkat menjadi 1420 kasus kecelakaan. Peningkatan jumlah kasus kecelakaan ini tentunya sangat merugikan berbagai pihak dan apabila tidak diketahui penyebab utamanya bisa berdampak fatal.

Data Kecelakaan tersebut akan sangat bermakna jika digali dengan tepat sehingga dapat menemukan pengetahuan dari data dan digunakan untuk mendapatkan informasi yang tersembunyi (Ogor, 2007). Dengan menerapkan teknik *Data Mining*, akan memecahkan masalah dengan menganalisis data yang dimiliki pihak kepolisian Sleman.

Data Mining adalah penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar. Terdapat tiga teknik Data Mining yaitu Asosiasi, Clustering dan Klasifikasi. Dalam penelitian ini penulis menggunakan teknik klasifikasi. “Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek” (Bustami, 2014). Salah satu teknik klasifikasi yaitu *Support Vector Machine* (SVM).

Support Vector Machine (SVM) dikembangkan oleh Boser, Guyon, Vapnik dan pertama kali dipresentasikan pada tahun 1992 di *Annual Workshop on Computational Learning Theory*. Metode ini merupakan metode mesin pembelajaran (*Learning Machine*) dengan tujuan menemukan fungsi pemisah (*hyperplane*) terbaik yang memisahkan dua buah kelas pada *input space* (Nugroho et al, 2003). Dewasa ini SVM telah berhasil diaplikasikan dalam problema dunia nyata (*real-world problems*), dan secara umum memberikan solusi yang lebih baik dibandingkan metode konvensional misalnya *Artificial Neural Network* (Nugroho, 2003). SVM juga bekerja dengan baik pada set data dengan dimensi yang tinggi.

“Dari latar belakang di atas, penulis tertarik meneliti tentang klasifikasi tingkat keparahan korban kecelakaan lalu lintas di Kabupaten Sleman tahun 2016 menggunakan metode Support Vector Machine.”

1.2 Rumusan Masalah

Dari latar belakang masalah yang telah dijelaskan, maka didapatkan rumusan masalah sebagai berikut :

1. Bagaimana gambaran umum kecelakaan lalu lintas Kabupaten Sleman, Yogyakarta tahun 2016-2017 ?
2. Bagaimana “hasil klasifikasi tingkat keparahan luka korban kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 menggunakan metode SVM (Support Vector Machine) *Polynomial Kernel SMOTE* ?”
3. Bagaimana tingkat akurasi hasil klasifikasi keparahan luka korban kecelakaan di Kabupaten Sleman, Yogyakarta tahun 2016-2017 dengan menggunakan metode *SVM (Support Vector Machine) Polynomial Kernel SMOTE* ?

1.3 Batasan Masalah

Agar penelitian ini tidak keluar dari pokok permasalahan yang dirumuskan, maka ruang lingkup pembahasan dibatasi yaitu pada data Kecelakaan Lalu Lintas di Kabupaten Sleman Yogyakarta tahun 2016-2017. Data diperoleh dari Polres Sleman Yogyakarta.

1.4 Tujuan Penelitian

1. Mengetahui gambaran umum kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta
2. Mengetahui “hasil klasifikasi tingkat keparahan luka korban kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 menggunakan metode SVM (Support Vector Machine) *Polynomial Kernel SMOTE*”.
3. Mengetahui tingkat akurasi hasil klasifikasi keparahan luka korban kecelakaan di Kabupaten Sleman, Yogyakarta tahun 2016-2017

dengan menggunakan metode *SVM (Support Vector Machine)*
Polynomial Kernel SMOTE

1.5 Manfaat Penelitian

1. Memberikan informasi pada pihak yang terkait dalam rangka mengurangi jumlah kecelakaan dan korban kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta.
2. Memberikan informasi tentang pola kecelakaan yang terjadi di Kabupaten Sleman, Yogyakarta melalui hasil klasifikasi data kecelakaan lalu lintas Kabupaten Sleman tahun 2016-2017

BAB II

TINJAUAN PUSTAKA

Penelitian terdahulu sebagai kajian untuk mengetahui hubungan antar penelitian yang dilakukan sebelumnya. Beberapa literature yang mendukung penelitian ini adalah sebagai berikut :

Penelitian oleh Octaviani, Wilandari dan Ispriyanti (2014) dengan judul penelitian “Penerapan Metode Klasifikasi Support Vector Machine (SVM) Pada Data Akreditasi Sekolah Dasar (SD) Di Kabupaten Magelang”. Penelitian ini membahas tentang akreditasi yang merupakan pengakuan institusi “pendidikan yang diberikan oleh pejabat yang berwenang, yaitu *Badan Akreditasi Nasional Sekolah / Madrasah (BAN-S / M)* setelah dinilai bahwa lembaga tersebut telah memenuhi delapan komponen penilaian akreditasi. Sekolah dasar, sebagai salah satu wajib belajar dasar, harus memiliki status akreditasi untuk menjamin kualitas pendidikan. Penelitian ini bertujuan untuk menerapkan metode klasifikasi *Support Vector Machine (SVM)* pada data akreditasi SD di Magelang. *Support Vector Machine (SVM)* adalah metode yang dapat digunakan sebagai klasifikasi prediktif dengan menggunakan konsep pencarian *hyperplane* (fungsi pemisah) yang dapat memisahkan data sesuai kelas. SVM menggunakan trik *kernel* untuk masalah *non linier* yang dapat mentransformasikan data menjadi ruang berdimensi tinggi dengan menggunakan fungsi *kernel*, sehingga data dapat diklasifikasikan secara linier. Hasil penelitian ini menunjukkan bahwa keakuratan prediksi klasifikasi SVM dengan menggunakan fungsi *kernel Gaussian RBF* adalah 93,902%. Ini dihitung dari 77 dari 82 sekolah dasar yang diklasifikasikan dengan benar dengan kelas aslinya”.

Penelitian oleh Athoillah, Irawan dan Imah (2015) dengan judul penelitian “Support Vector Machine Untuk Image Retrieval”. Penelitian ini membahas tentang Gerbang Tol Otomatis (GTO) yang mulai diberlakukan di beberapa pintu masuk jalan tol di Indonesia, dengan sistem seperti ini tentunya diperlukan sebuah alat yang dapat memilah atau membedakan jenis kendaraan apa yang ada didepan gerbang karena tidak semua jenis kendaraan boleh masuk jalan tol, permasalahan

membedakan objek tersebut pada dasarnya adalah masalah klasifikasi. Salah satu algoritma yang baik dalam memecahkan masalah pengklasifikasian adalah Support Vector Machine (SVM), Penelitian yang diusulkan dalam paper ini adalah untuk membangun sebuah metode klasifikasi dengan algoritma SVM yang di ujicobakan dalam image retrieval. Dari hasil yang diperoleh, menunjukkan bahwa metode yang diajukan memiliki rata-rata akurasi sebesar 82,22%, presisi 82,82% dan recall 82,22% untuk dataset citra sebanyak 1000 dengan 2 kategori kelas citra dengan objek roda dua dan roda empat, hasil penelitian juga menunjukkan bahwa rata-rata waktu komputasi selama proses training 19,50 detik sedangkan proses testing rata-rata membutuhkan waktu 9,47 detik

Penelitian oleh Ispriyanti dan Hoyyi (2016) dengan judul penelitian “Analisis Klasifikasi Masa Studi Mahasiswa Prodi Statistika UNDIP dengan Metode *Support Vector Machine* (SVM) dan ID3 (*Iterative Dichotomiser 3*)”. Penelitian ini membahas mengenai wisuda merupakan tahap akhir dari proses kegiatan belajar di perguruan tinggi. Masa studi akademik UNDIP dalam 8 semester (4 tahun) atau kurang dan maksimal 14 semester (7 tahun). Jurusan Statistik adalah salah satu dari enam jurusan di Fakultas Sains dan Matematika UNDIP. Masa studi di jurusan ini dapat dipengaruhi oleh banyak faktor. Faktor tersebut adalah *Grade Point Average* (GPA) atau IPK, jalur jender, beasiswa, *parttime*, organisasi, dan masuk universitas. Tujuan dari makalah ini adalah untuk menentukan klasifikasi faktor akurasi. Dengan menggunakan SVM (*Support Vector Machine metode*) dan ID3 (*Iterative Dichotomiser 3*). Perbandingan SVM dan metode ID3, baik untuk pelatihan dan pengujian data menghasilkan akurasi yang baik, yaitu 90%. Terutama data training ID3 memberikan hasil yang lebih baik dari pada SVM.

Penelitian oleh Dwi Putri (2017) dengan judul penelitian “Klasifikasi Status Kelulusan Mahasiswa Menggunakan Metode Support Vector Machine”. Penelitian ini membahas tentang kelulusan yang merupakan hasil akhir pencapaian yang membanggakan dalam menempuh suatu pendidikan pada jenjang tertentu. Masa studi mahasiswa jenjang S1 secara normal menurut kurikulum yang berlaku adalah 8 semester atau 4 tahun. Mahasiswa dinyatakan lulus tepat waktu apabila

menyelesaikan masa studi kurang dari sama dengan 4 tahun. Tujuan dari penelitian ini untuk mengetahui hasil klasifikasi yang tepat antara fungsi Kernel Linear, Polynomial dan Radial Basis Function (RBF) terhadap status kelulusan mahasiswa jurusan Statistika, Farmasi dan Ilmu Kimia UII. Dengan menggunakan metode Support Vector Machine (SVM) dengan tiga fungsi Kernel yaitu Linear, Polynomial dan Radial Basis Function (RBF). Variabel yang digunakan dalam penelitian ini yaitu Jenis Kelamin, Jenis SMA, Jurusan SMA, Asal Daerah, IPK, Pekerjaan Orang Tua dan Lama Studi. Dengan menggunakan metode SVM dengan tiga fungsi Kernel didapatkan hasil bahwa fungsi Kernel Polynomial lebih baik dibandingkan metode SVM dengan fungsi Kernel Linear dan RBF untuk data tersebut dengan tingkat keakuratan klasifikasi sebesar 71.37%

BAB III

LANDASAN TEORI

3.1. Kecelakaan Lalu Lintas

Menurut “Undang-undang Nomor 22 tahun 2009 tentang kecelakaan lalu lintas dan angkutan jalan, disebutkan kecelakaan lalu lintas adalah suatu peristiwa di jalan yang tidak diduga dan tidak disengaja melibatkan kendaraan dengan atau tanpa pengguna jalan lain yang mengakibatkan kerugian dan/atau kerugian harta benda. Sedangkan menurut Kamus Besar Bahasa Indonesia pengertian kecelakaan adalah kejadian (peristiwa) yang menyebabkan orang celaka”.

3.1.1 Karakteristik Kecelakaan Lalu Lintas

“Berdasarkan Undang-undang Nomor 22 tahun 2009 pasal 229, karakteristik kecelakaan lalu lintas dapat dibagi menjadi 3 golongan, yaitu:

1. “Kecelakaan lalu lintas ringan, yaitu kecelakaan yang mengakibatkan kerusakan kendaraan dan/atau barang”.
2. “Kecelakaan lalu lintas sedang, yaitu kecelakaan yang mengakibatkan luka ringan dan kerusakan kendaraan dan/atau barang”.
3. “Kecelakaan lalu lintas berat, yaitu kecelakaan yang mengakibatkan korban meninggal dunia atau luka berat”

“Pada suatu kecelakaan lalu lintas yang terjadi, ada beberapa kriteria keparahan korban kecelakaan menurut PP Nomor 43 Tahun 1993 Pasal 93, antara lain sebagai berikut”.

1. Korban Meninggal

“Korban meninggal adalah korban yang dipastikan meninggal dunia sebagai akibat kecelakaan lalu lintas dalam jangka waktu paling lama 30 hari setelah kecelakaan tersebut”.

2. Korban Luka Berat

“Korban luka berat adalah korban yang karena luka-lukanya menderita cacat tetap atau harus dirawat dalam jangka waktu lebih dari 30 hari sejak terjadi kecelakaan”.

3. Korban Luka Ringan

“Korban luka ringan adalah korban yang tidak termasuk dalam kategori korban meninggal dan korban luka berat”.

3.1.2 “Faktor-Faktor Penyebab Kecelakaan lalu Lintas”

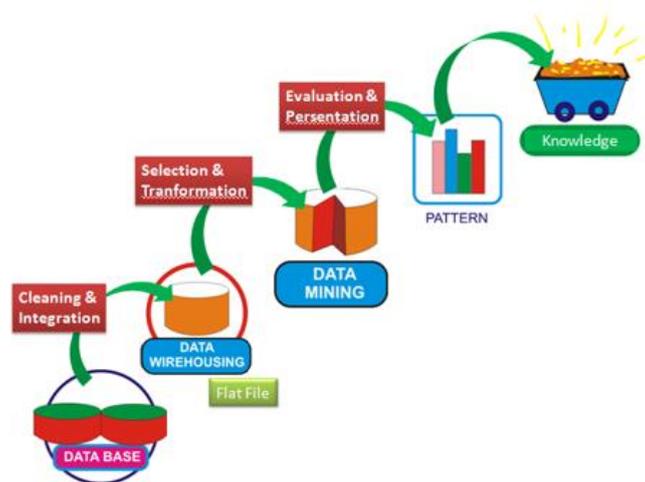
Ada banyak faktor yang menyebabkan kecelakaan lalu lintas. Kecelakaan lalu lintas ke dalam tiga komponen utama sistem lalu lintas jalan (Azizirrahman, 2015) :

1. Manusia, faktor yang dapat dilihat dari pengguna kendaraan bermotor, yaitu: kecepatan tinggi/ ugal-ugalan, muatan yang berlebihan dan tidak tertib (tidak memakai helm dan melanggar rambu lalu lintas dan marka jalan).
2. Kendaraan, faktor yang dapat dilihat dari kendaraan, yaitu: kaca spion dan lampu kendaraan (lampu utama, lampu indikator/sein dan lampu rem).
3. Faktor Lingkungan Fisik Jalan, faktor yang dapat dilihat dari jalan, yaitu: jalan rusak, jalan berlubang, jalan tergenang, jalan gelap, tanpa marka/rambu jalan dan tikungan tajam.

3.2 Data Mining

“*Data mining* adalah suatu istilah yang digunakan untuk menemukan pengetahuan yang tersembunyi di dalam *database*. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang bermanfaat yang tersimpan di dalam *database* besar (Turban et al, 2005). Menurut Gartner Group *Data Mining* adalah 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).

Data mining adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran/berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan. Dimana langkah-langkah untuk melakukan *Data Mining* adalah sebagai berikut :



(Sumber: Hadinata. 2013)

Gambar 3.1. Langkah-langkah untuk melakukan *Data Mining*

- 1) “*Data cleaning* (untuk menghilangkan *noise* data yang tidak konsisten) *Data integration* (di mana sumber data yang terpecah dapat disatukan)”.
- 2) “*Data selection* (di mana data yang relevan dengan tugas analisis dikembalikan ke dalam *database*)”.

- 3) “*Data transformation* (di mana data berubah atau bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa)”.
- 4) “*Data mining* (proses esensial di mana metode yang intelejen digunakan untuk mengekstrak pola data)”.
- 5) “*Pattern evolution* (untuk mengidentifikasi pola yang benar-benar menarik yang mewakili pengetahuan berdasarkan atas beberapa tindakan yang menarik)”.
- 6) “*Knowledge presentation* (di mana gambaran teknik visualisasi dan pengetahuan digunakan untuk memberikan pengetahuan yang telah ditambang pada *user*)”.

“Kegunaan *Data Mining* adalah untuk menspesifikasikan pola yang harus ditemukan dalam tugas *Data Mining*. Secara umum tugas *Data Mining* dapat diklasifikasikan ke dalam dua kategori: deskriptif dan prediktif. Tugas menambang secara deskriptif adalah untuk mengklasifikasikan sifat umum suatu data di dalam *database*. Tugas *Data Mining* secara prediktif adalah untuk mengambil kesimpulan terhadap data terakhir untuk membuat prediksi”.

“Terdapat beberapa teknik *Data Mining* yang sering digunakan dalam literatur. Namun ada 3 teknik *data mining* yang populer” (Haryati et al, 2015), yaitu:

1. *Association Rule Mining*

“*Association Rule Mining* adalah teknik mining untuk menemukan asosiatif antara kombinasi atribut. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan dapat mengatur penempatan barangnya atau merancang strategi pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu”.

2. *Clustering*

“Berbeda dengan *Association Rule Mining* dan klasifikasi dimana kelas data telah ditentukan sebelumnya dapat dipakai untuk memberikan label pada kelas data yang belum diketahui. Karena itu *Clustering* sering digolongkan sebagai metode *unsupervised learning*. Prinsip *Clustering* adalah memaksimalkan kesamaan antar *Cluster*. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi”.

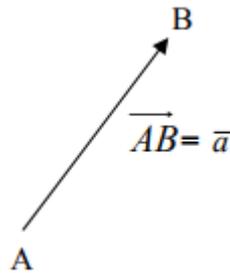
3. Klasifikasi

“Suatu proses yang menemukan properti-properti yang sama pada sebuah himpunan obyek di dalam sebuah basis data, dan mengklasifikasikannya ke dalam kelas-kelas yang berbeda menurut model klasifikasi yang ditetapkan”.

“Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah”.

3.3 Vektor

“Vektor adalah suatu besaran yang mempunyai besar dan arah. Secara grafis suatu vektor ditunjukkan sebagai potongan garis yang mempunyai arah. Besar atau kecilnya vektor ditentukan oleh panjang atau pendeknya potongan garis. Sedangkan arah vektor ditunjukkan dengan tanda anak panah” (Anonim, 2017a).



Gambar 3.2. Vektor

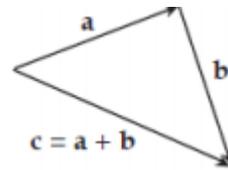
Dalam **gambar 3.2**, titik A disebut titik awal (*initial point*) dan titik P disebut titik terminal (*terminal point*). “Contoh vektor misalnya lintasan, kecepatan, percepatan, dan gaya. Skalar adalah suatu kuantita yang mempunyai besaran tetapi tidak mempunyai arah. Suatu skalar adalah bilangan nyata dan secara simbolik dapat ditulis dengan huruf kecil. Operasi skalar mengikuti aturan yang sama dengan aturan operasi aljabar elementer” (Anonim, 2017b).

Secara geometri penjumlahan vektor $\vec{a} \pm \vec{b} = \vec{c}$ dapat dilakukan dengan dua cara yaitu sebagai berikut :

a. Cara Segitiga

Titik pangkal vektor \vec{b} berimpit ruas dengan titik ujung vektor \vec{a} . Jumlah vektor \vec{a} dan \vec{b} didapat dengan menarik ruas garis dari titik pangkal

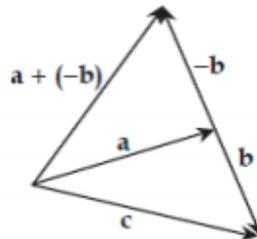
vektor \vec{a} ke titik ujung vektor \vec{b} . Ruas garis bawah ini diwakili oleh vektor \vec{c} . Sehingga $\vec{a} \pm \vec{b} = \vec{c}$.



Gambar 3.3. Cara Segitiga

b. Aturan Jajar Genjang

Titik pangkal vektor \vec{a} dan \vec{b} harus berimpit.



Gambar 3.4. Aturan Jajar Genjang

Perkalian Vektor

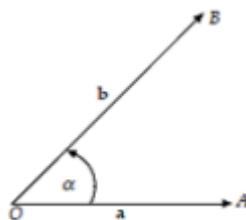
a. Perkalian skalar dengan vektor

Jika k skalar tak nol dan vektor $\vec{a} = a_1\vec{i} + a_2\vec{j} + a_3\vec{k}$ maka vektor $k\vec{a} = (ka_1, ka_2, ka_3)$.

b. Perkalian skalar dua vektor

Jika vektor $\vec{a} = a_1\vec{i} + a_2\vec{j} + a_3\vec{k}$ dan vektor $\vec{b} = b_1\vec{i} + b_2\vec{j} + b_3\vec{k}$ maka $\vec{a} \cdot \vec{b} = a_1b_1 + a_2b_2 + a_3b_3$.

c. Perkalian skalar dua vektor jika membentuk sudut



Gambar 3.5. Perkalian skalar dua vektor jika membentuk sudut

Jika \vec{a} dan \vec{b} vektor tak nol dan sudut α diantara vektor \vec{a} dan \vec{b} . Maka perkalian skalar vektor \vec{a} dan \vec{b} adalah $=|\vec{a}||\vec{b}|\cos\alpha$. Jika \vec{a} , \vec{b} dan \vec{c} vektor di \mathbb{R}^2 dan \mathbb{R}^3 dan k, l adalah scalar tak nol maka berlaku aturan vektor berikut :

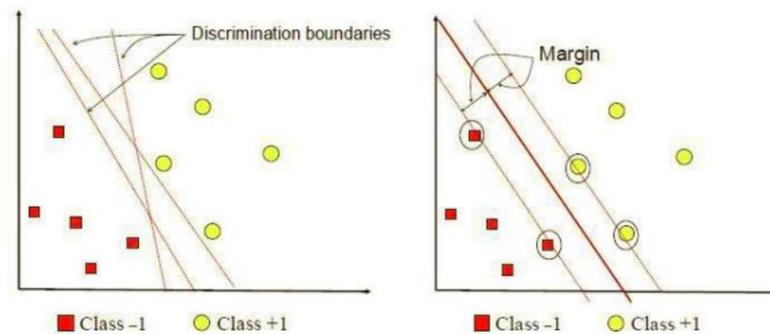
1. $\vec{a} + \vec{b} = \vec{b} + \vec{a}$
2. $(\vec{a} + \vec{b}) + \vec{c} = \vec{a} + (\vec{b} + \vec{c})$
3. $\vec{a} + \vec{0} = \vec{0} + \vec{a} = \vec{a}$
4. $\vec{a} + (-\vec{a}) = \vec{0}$
5. $\vec{a} = \vec{a}$
6. $k(l\vec{a}) = (kl)\vec{a}$
7. $k(\vec{a} + \vec{b}) = k\vec{a} + k\vec{b}$
8. $(k + l)\vec{a} = k\vec{a} + l\vec{a}$
9. $\vec{a} \cdot \vec{b} = \vec{b} \cdot \vec{a}$
10. $k(\vec{a} \cdot \vec{b}) = (k\vec{a}) \cdot \vec{b} = \vec{a} \cdot (k\vec{b})$
11. $\vec{a} \cdot (\vec{b} + \vec{c}) = \vec{a} \cdot \vec{b} + \vec{a} \cdot \vec{c}$
12. $\vec{a} \cdot \vec{a} = |\vec{a}|^2$

3.4. Support Vector Machine (SVM)

SVM adalah kumpulan metode pembelajaran yang digunakan untuk klasifikasi dan regresi yang termasuk dalam bagian klasifikasi linier umum. Properti khusus SVM adalah meminimalkan kesalahan klasifikasi empiris dan memaksimalkan margin geometrik. Jadi SVM disebut *Maximum Margin Classifiers*. SVM didasarkan pada Structural Risk Minimization (SRM). Vektor input peta SVM ke ruang dimensi yang lebih tinggi dimana medan *hyperplane* separasi maksimal dibuat. Dua *hyperplane* paralel dibangun di setiap sisi *hyperplane* yang memisahkan data. Diasumsikan bahwa semakin besar margin atau jarak antara hyperplanes, semakin baik generalisasi kesalahan pengklasifikasian (Durgesh, 2009)

3.4.1. SVM Pada Data Terpisah Secara *Linear*

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari *hyperplane* terbaik yang berfungsi sebagai pemisah dua buah *class* pada input space.



(Sumber: Riyandi. 2015)

Gambar 3.6. Ilustrasi SVM menemukan *hyperplane* terbaik yang memisahkan dua kelas -1 dan +1

Gambar 3.6 memperlihatkan anggota dari dua buah kelas : +1 dan -1. Pada kelas -1 disimbolkan dengan kotak berwarna merah sementara kelas +1 disimbolkan dengan lingkaran berwarna kuning. “Masalah klasifikasi dapat diterjemahkan dengan usaha menemukan *hyperplane* yang memisahkan antara kedua kelompok tersebut. *Hyperplane* pemisah terbaik antara kedua kelas dapat ditemukan dengan mengukur *margin hyperplane* dengan titik terdekat masing-masing kelas”. Titik yang paling dekat ini disebut dengan *support vector*. Garis solid pada gambar 3.6 sebelah kanan menunjukkan *hyperplane* terbaik, yaitu yang terletak pada tengah-tengah kedua *class*, sedangkan titik merah dan kuning yang berada dalam lingkaran hitam adalah *support vector*. Label masing-masing dinotasikan $y_i \in \{-1, +1\}$ untuk $i = 1, 2, 3, \dots, l$. *Hyperplane* klasifikasi linear SVM seperti pada persamaan 3.1.

$$\vec{w} \cdot \vec{x} + b = 0 \quad (3.1)$$

Keterangan:

\vec{w} = *vector* bobot

\vec{x} = nilai masukan atribut

b = bias

\vec{x}_i yang termasuk kelas -1 (sampel negatif) dapat dirumuskan memenuhi pertidaksamaan :

$$\vec{w} \cdot \vec{x} \leq -1 \quad (3.2)$$

Sedangkan \vec{x}_i yang termasuk kelas +1 (sampel positif)

$$\vec{w} \cdot \vec{x} + b \geq +1 \quad (3.3)$$

Margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara jarak dan titik terdekatnya, yaitu $1/\|\vec{w}\|$. Hal ini dapat dirumuskan sebagai *Quadratic Programming (QP) problem*, yaitu mencari titik minimal persamaan 3.4, dengan memperhatikan *constraint* persamaan 3.5.

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 \quad (3.4)$$

$$y_i (\vec{x}_i \cdot \vec{w} + b) - 1 \geq 0, \forall_i \quad (3.5)$$

“ x_i adalah data input, adalah keluaran dari data x_i, w, b adalah parameter-parameter yang di cari nilainya. Dalam formulasi di atas, ingin meminimalkan fungsi tujuan (*obyektif function*) $\frac{1}{2} \|\vec{w}\|^2$ atau memaksimalkan kuantitas $\|\vec{w}\|^2$ dengan memperhatikan pembatas $y_i (wx_i + b) \geq 1$. Bila output data $y_i = +1$, maka pembatas menjadi $(wx_i + b) \geq 1$. Sebaliknya bila $y_i = -1$, pembatas menjadi $(wx_i + b) \leq -1$. Di dalam kasus yang tidak feasible (infeasible) dimana beberapa data mungkin tidak bisa dikelompokkan secara benar, formulasi matematikanya menjadi berikut”.

$$\min_{\vec{w}} \tau(w) = \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^l t_i \quad (3.6)$$

$$y_i (\vec{x}_i \cdot \vec{w} + b) + t_i \geq 1, \forall_i \quad (3.7)$$

$$t_i \geq 0, i = 1, \dots, l$$

Problem ini dapat dipecahkan dengan berbagai teknik komputasi, di antaranya *Lagrange Multiplier*.

$$L(w, b, \alpha) = \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^l \alpha_i (y_i ((\bar{x}_i \cdot \bar{w} + b) - 1)) \quad (3.8)$$

$$L(w, b, \alpha) = \frac{1}{2} \|\bar{w}\|^2 - \sum_{i=1}^l \alpha_i y_i (x_i \cdot \bar{w}) - b \sum_{i=1}^l \alpha_i y_i + \sum_{i=1}^l \alpha_i \quad (3.9)$$

Dengan tambahan kosntrein, $\alpha_i \geq 0$ (nilai dari koefisien *lagrange*). Dengan meminimumkan L terhadap w dan b.

$$\frac{\partial}{\partial b} L(w, b, \alpha) = 0 \quad (3.10)$$

$$\frac{\partial}{\partial w} L(w, b, \alpha) = 0 \quad (3.11)$$

Dari persamaan 3.8 dan persamaan 3.9 diperoleh persamaan berikut :

$$\sum_{i=1}^l \alpha_i y_i = 0 \quad (3.12)$$

$$w = \sum_{i=1}^l \alpha_i y_i x_i \quad (3.13)$$

“Vektor w sering kali bernilai besar (mungkin tak terhingga), tetapi nilai α_i terhingga. Untuk itu, formula *lagrangian* L_p (*primal problem*) diubah kedalam *dual problem*. Dengan mensubsitusikan persamaan 3.13 ke LP diperoleh *dual problem* L_d dengan konstrain berbeda”.

$$L_d(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j \quad (3.14)$$

“Dimana α_i adalah *Lagrange Multiplier*, yang bernilai 0 atau positif ($\alpha_i \geq 0$) . Nilai optimal dari persamaan 3.4 dapat dihitung dengan meminimumkan L terhadap \bar{w} dan b, dan memaksimalkan L terhadap α_i , dengan memperhatikan sifat bahwa pada titik optimal *gradient* $L = 0$ persamaan 3.6 dapat dimodifikasi sebagai

maksimalisasi problem yang hanya mengandung α_i , sebagaimana terlihat pada persamaan 3.15 dan 3.16 dibawah ini”.

$$\max_{\alpha} L_d = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j \bar{x}_i \bar{x}_j \quad (3.15)$$

$$\alpha_i \geq 0 (i = 1, 2, \dots, l) \quad \sum_{i=1}^l \alpha_i y_i = 0 \quad (3.16)$$

Dengan demikian, maka akan diperoleh α_i yang kebanyakan bernilai positif yang disebut sebagai *support vector*.

3.4.2. SVM Pada Data Tidak Terpisah Secara Linear

“Kasus data yang tidak terpisah secara linear diasumsikan bahwa kelas pada *input space* tidak dapat terpisah secara sempurna. Hal ini menyebabkan pada persamaan 3.5 tidak dapat terpenuhi, sehingga optimisasi tidak dapat dilakukan, untuk mengatasi masalah ini SVM dirumuskan ulang dengan memperkenalkan teknik *softmargin*. Dalam *softmargin* persamaan 3.5 dimodifikasi dengan menggunakan *slack* variabel sehingga terlihat pada persamaan 3.17”.

$$y_i (\bar{x}_i \cdot \vec{w} + b) \geq 1 - \xi_i, \quad \forall_i \quad (3.17)$$

Dengan demikian, persamaan 3.4 diubah menjadi persamaan 3.18.

$$\min_{\vec{w}} \tau(w, \xi) = \frac{1}{2} \|\vec{w}\|^2 + C \sum_{i=1}^l \xi_i \quad (3.18)$$

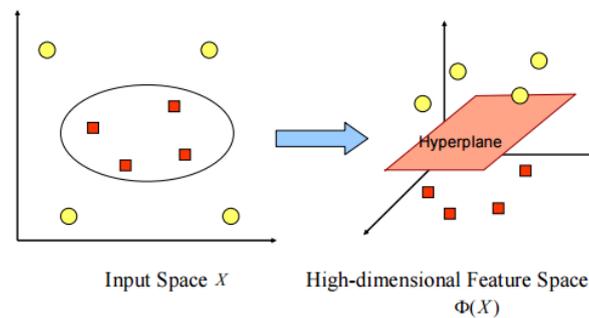
“Parameter C digunakan untuk mengontrol *tradeoff* antara *margin* dan kesalahan klasifikasi ξ . Nilai C yang besar akan memeberikan pelanggaran yang lebih besar terhadap *error* klasifikasi tersebut (Sembiring, 2007)”.

3.4.3. Kernel Trick dan Non-Linear Classification pada SVM

“Pada umumnya masalah dalam domain dunia nyata (*real-world problem*) jarang yang bersifat *linear separable*. Kebanyakan bersifat *non-linear*. Untuk menyelesaikan masalah *non-linear*, SVM dimodifikasi dengan memasukkan fungsi *Kernel*. Dalam *non-linear* SVM, pertama-tama data x dipetakan oleh fungsi $\Phi(x)$ ke ruang vektor yang berdimensi lebih tinggi”. Pada ruang vektor yang baru ini,

hyperplane yang memisahkan kedua kelas tersebut dapat dikonstruksikan. Hal ini sejalan dengan teori Cover yang menyatakan “*Jika suatu transformasi bersifat non-linear dan dimensi dari feature space cukup tinggi, maka data pada input space dapat dipetakan ke feature space yang baru, dimana pattern-pattern tersebut pada probabilitas tinggi dapat dipisahkan secara linear*” (Susilowati et al, 2015).

“Ilustrasi dari konsep ini dapat dilihat pada gambar 3.7. Pada gambar 3.7 diperlihatkan data pada kelas kuning dan data pada kelas merah yang berada pada *input space* berdimensi dua tidak dapat dipisahkan secara *linear*. Selanjutnya fungsi Φ memetakan tiap data pada *input space* tersebut ke ruang vektor baru yang berdimensi lebih tinggi (dimensi 3), dimana kedua kelas dapat dipisahkan secara *linear* oleh sebuah *hyperplane*”.



(Sumber: Riyandi, 2015)

Gambar 3.7. Fungsi Φ memetakan data ke ruang vector yang berdimensi lebih tinggi, sehingga kedua kelas dapat dipisahkan secara *linear* oleh sebuah *hyperplane*

Selanjutnya proses pembelajaran pada SVM dalam menemukan titik-titik *support vector*, hanya bergantung pada *dot product* dari data yang sudah ditransformasikan pada ruang baru yang berdimensi lebih tinggi, yaitu $\Phi(x_i) \cdot \Phi(x_j)$. Karena umumnya transformasi Φ ini tidak diketahui, dan sangat sulit untuk difahami secara mudah, maka perhitungan *dot product* dapat digantikan dengan fungsi *Kernel Trick* $K(\vec{x}_i, \vec{x}_j)$ yang terlihat pada persamaan 3.19.

$$\begin{aligned}
K(\vec{x}_i, \vec{x}_j) &= \Phi(\vec{x}_i) \cdot \Phi(\vec{x}_j) \\
f(\Phi(\vec{x})) &= \vec{w} \cdot \Phi(\vec{x}) + b \\
&= \sum_{i=1, x_i \in SV}^n \alpha_i y_i \Phi(\vec{x}) \cdot \Phi(\vec{x}_i) + b \\
&= \sum_{i=1, x_i \in SV}^n \alpha_i y_i K(x, x_i) + b
\end{aligned} \tag{3.19}$$

Beberapa *Kernel* yang umum dipakai pada SVM adalah:

a. *Linear*

“*Kernel Trick Linear* ini dinyatakan dalam persamaan”.

$$K(\vec{x}_i, \vec{x}_j) = \vec{x}_i \cdot \vec{x}_j \tag{3.20}$$

b. *Polynomial*

“*Kernel Trick Polynomial* cocok digunakan untuk menyelesaikan masalah klasifikasi, dimana dataset pelatihan sudah normal. *Kernel Trick* ini dinyatakan dalam persamaan”.

$$K(\vec{x}_i, \vec{x}_j) = (\vec{x}_i \cdot \vec{x}_j + 1)^p \tag{3.21}$$

X_i dan X_j adalah pasangan dua data training. Angka 1 merupakan nilai cost yang merupakan konstanta.

c. *Radial Basis Function* atau *Gaussian*

“*Kernel Trick Radial Basis Function* atau *Gaussian* merupakan *Kernel* yang paling banyak digunakan untuk menyelesaikan masalah klasifikasi untuk dataset yang tidak terpisah secara *linear*, dikarenakan akurasi pelatihan dan akurasi prediksi yang sangat baik pada *Kernel* ini, dimana *Kernel Radial Basis Function* dinyatakan dalam persamaan 3.22”.

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\gamma \|\vec{x}_i - \vec{x}_j\|^2\right). \tag{3.22}$$

“Pada awalnya SVM dikembangkan untuk persoalan klasifikasi dua kelas, kemudian dikembangkan kembali untuk klasifikasi multikelas (Santosa, 2007). Dalam klasifikasi kasus multikelas, hyperplane yang terbentuk adalah lebih dari satu. Salah satu metode pendekatan yang digunakan adalah satu lawan semua (SLA). Metode SLA untuk kasus klasifikasi k-kelas, menemukan k hyperplane dimana k adalah banyak kelas

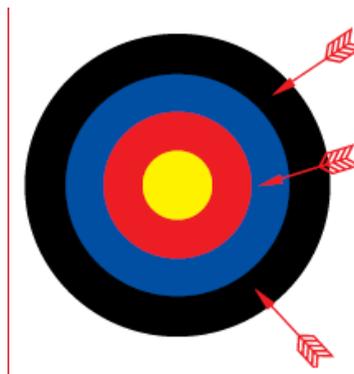
dan ρ adalah hyperplane. Dalam metode ini $\rho(\ell)$ diujikan dengan semua data dari kelas ℓ dengan label +1, dan semua data dari kelas lain dengan label -1”.

“Konsep pada SLA yaitu dimisalkan pada kasus tiga kelas, kelas 1, 2, dan 3. Bila akan diujikan $\rho(1)$, semua data dalam kelas 1 diberi label +1 dan data dari kelas 2 dan 3 diberi label -1. Pada $\rho(2)$, semua data dalam kelas 2 diberi label +1 dan data dari kelas 1 dan 3 diberi label -1. Begitu juga untuk $\rho(3)$, semua data dalam kelas 3 diberi label +1 dan data dari kelas 1 dan 2 diberi label -1. Kemudian dicari hyperplane dengan algoritma SVM dua kelas. Maka akan didapat hyperplane untuk masing-masing kelas diatas. Kemudian kelas dari suatu data baru x menurut Hsu dan Lin (2002) ditentukan berdasarkan nilai terbesar dari hyperplane”.

3.5. Accuracy dan Precision

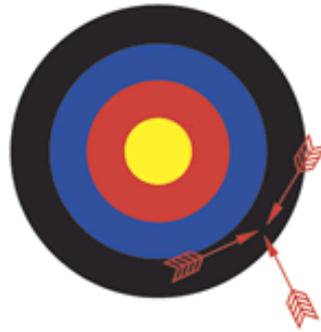
Secara umum, ukuran *accuracy* dalam klasifikasi yaitu mengukur rasio prediksi yang benar dari jumlah total contoh yang dievaluasi, sedangkan *precision* mengukur pola positif yang terprediksi benar (*True Positive*) dari total pola yang terprediksi dalam kelas positif (Hossin, 2015).

Perbedaan antara akurasi dan presisi diilustrasikan di bawah. Sasaran anak panah yang di target mewakili nilai sebenarnya dari sebuah pengukuran.



(Sumber : National Physical Laboratory, 2015)

Gambar 3.8 Akurasi rendah, presisi rendah,



(Sumber : National Physical Laboratory, 2015)

Gambar 3.9 Akurasi rendah, presisi tinggi



(Sumber : National Physical Laboratory, 2015)

Gambar 3.10 Akurasi tinggi, presisi rendah



(Sumber : National Physical Laboratory, 2015)

Gambar 3.11 Akurasi tinggi, presisi tinggi

3.6. *Imbalance Dataset*

Suatu class pada dataset dengan pendistribusian class yang tidak seimbang (*imbalance class*) menimbulkan kejadian klasifikasi lebih condong ke class mayoritas dibandingkan dengan class minoritas. Ketidakseimbangan class pada sebuah dataset merupakan suatu permasalahan dalam *machine learning*, dimana jumlah class mayoritas (*negative*) lebih besar dari pada jumlah class minoritas

(*positive*). Dengan kata lain, jumlah class negative (mayoritas) lebih besar jumlahnya dibandingkan dengan class positive (minoritas). Permasalahan class imbalance merupakan sebuah permasalahan yang sudah lazim ditemukan pada dataset di berbagai bidang, termasuk prediksi cacat software, deteksi tumpahan minyak dari citra satelit, deteksi penipuan kartu kredit online, dan diagnosis penyakit (Phoungphol, 2013). Untuk mengatasi ketidakseimbangan kelas adalah dengan melakukan oversampling. Oversampling merupakan mekanisme untuk menyeimbangkan distribusi kelas dengan cara replikasi instance kelas minoritas secara acak. Namun kekurangan dari oversampling adalah peningkatan kemungkinan munculnya overfitting karena mekanisme ini membuat duplikasi instance secara persis. Solusi yang dianggap tepat dalam menangani kasus overfitting pada oversampling yaitu dilakukan suatu pendekatan *rebalancing* pada dataset dengan menggunakan “*state of the art*” metode *sampling* yaitu *SMOTE* (*Synthetic Minority Over-sampling Technique*) (Fernandy, 2009).

3.6.1 SMOTE (*Synthetic Minority Over-sampling Technique*)

“SMOTE merupakan teknik oversampling yang baik dan efektif untuk menangani *overfitting* pada proses *oversampling* dalam menangani ketidakseimbangan di kelas modul yang cacat pada kelas minoritas (positif). Teknik SMOTE dipilih karena efektivitas dalam penanganan masalah ketidakseimbangan kelas dalam dataset cacat perangkat lunak” (Riquelme, 2008).

“Metode SMOTE menambah jumlah data kelas minor agar setara dengan kelas mayor dengan cara membangkitkan data buatan. Tujuan penambahan data ini agar jumlah data minor setara dengan data mayor. Data buatan atau sintesis tersebut dibuat berdasarkan *k*-tetangga terdekat (*k-nearest neighbor*). Jumlah *k*-tetangga terdekat ditentukan dengan mempertimbangkan kemudahan dalam melaksanakannya. Pembangkitan data buatan yang berskala numerik berbeda dengan kategorik”.

“Data numerik diukur jarak kedekatannya dengan jarak Euclidean sedangkan data kategorik lebih sederhana yaitu dengan nilai modus.

Perhitungan jarak antar contoh kelas minor yang peubahnya berskala kategorik dilakukan dengan rumus *Value Difference Metric* (VDM) yaitu” :

$$\Delta(X, Y) = W_x W_y \sum_{i=1}^N \delta(x_i, y_i)^r \quad (3.23)$$

Dimana :

$\Delta(X, Y)$ = jarak antara amatan X dengan Y

W_x, w_y = bobot amatan (dapat diabaikan)

N = banyaknya peubah penjelas

R = bernilai 1 (jarak Manhattan)
atau 2 (jarak Euclidean)

$\Delta(x_i, y_i)^r$ = jarak antar kategori, dengan rumus :

$$\Delta(V_1, V_2) = \sum_{i=1}^n \left| \frac{c_{1i}}{c_1} - \frac{c_{2i}}{c_2} \right| \quad (3.24)$$

Dimana :

$\Delta(V_1, V_2)$ = jarak antara nilai V1 dan V2

C_{1i} = banyaknya V1 yang termasuk kelas i

C_{2i} = banyaknya V2 yang termasuk kelas i

I = banyaknya kelas; $i = 1, 2, \dots, m$

C_1 = banyaknya nilai 1 terjadi

C_2 = banyaknya nilai 2 terjadi

N = banyaknya kategori

K = konstanta (biasanya 1)

Prosedur pembangkitan data buatan untuk :

1. Data Numerik

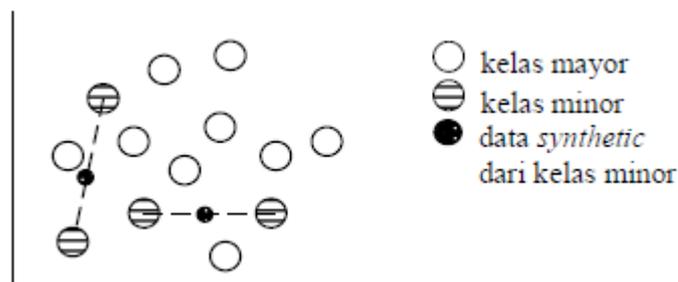
- a. Hitung perbedaan antar vektor utama dengan k -tetangga terdekatnya.
- b. Kalikan perbedaan dengan angka yang diacak di antara 0 dan 1.

- c. Tambahkan perbedaan tersebut ke dalam nilai utama pada vektor utama asal sehingga diperoleh vektor utama baru.

2. Data Kategorik

- a. Pilih mayoritas antara vektor utama yang dipertimbangkan dengan k -tetangga terdekatnya untuk nilai nominal. Jika terjadi nilai sama maka pilih secara acak.
- b. Jadikan nilai tersebut data contoh kelas buatan baru.

Ilustrasi distribusi data setelah diterapkan metode SMOTE dapat dilihat pada **Gambar 3.11**.



(Sumber : Angelina, 2010)

Gambar 3.12 Ilustrasi SMOTE

Meskipun SMOTE cukup efektif dalam meningkatkan akurasi klasifikasi dari data minoritas, tetapi masih ada masalah, antara lain, yaitu terjadinya overgeneralisasi. Data hasil sintesis SMOTE masih memungkinkan untuk menyebar pada data minoritas dan mayoritas, sehingga akan mengurangi kinerja klasifikasi (Zheng, 2015).

3.7. Confusion Matrix

Dalam menguji keefektifan suatu klasifikasi dibutuhkan suatu pengukuran evaluasi. Pengukuran tersebut didapatkan dalam sebuah *set confusion matrix* (Chin, 2010). “*Confusion matrix* merupakan sebuah tabel 3.1 yang terdiri atas banyaknya baris data uji yang diprediksi benar dan tidak benar oleh model klasifikasi, digunakan untuk menentukan kinerja suatu model klasifikasi”.

Tabel 3.1. *Confusion Matrix*

Predicted class	Actual Class		
		Positif	Negatif
Positif	TP	FN	
Negatif	FP	TN	

Keterangan :

TP (True Positif) adalah kelas yang diprediksi positif dan benar.

TN (True Negatif) adalah kelas yang diprediksi negatif dan benar.

FP (False Positif) adalah kelas yang diprediksi positif dan salah.

FN (False Negatif) adalah kelas yang diprediksi negatif dan salah.

Sehingga akurasi dari klasifikasi dapat diperoleh dari penjumlahan true positif dan true negatif dibagi total untuk melihat kinerja secara keseluruhan dengan rumus berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.25)$$

Sedangkan *error rate* dari klasifikasi dapat diperoleh dari penjumlahan False negatif dan false positif dibagi total berikut matematisnya :

$$Error = \frac{FN + FP}{TP + TN + FP + FN} \quad (3.26)$$

Presisi diperoleh dari true positif dibagi penjumlahan dari true positif dan false positif :

$$Presisi = \frac{TP}{TP + FP} \quad (3.27)$$

BAB IV

METODE PENELITIAN

4.1 Populasi dan Sampel Penelitian

Penelitian ini menggunakan populasi seluruh kecelakaan lalu lintas di Kabupaten Sleman . Sampel penelitian ini menggunakan data kecelakaan lalu lintas Kabupaten Sleman tahun 2016-2017.

4.2 Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder dari Polres Kabupaten Sleman, DIY yang diakses pada bulan Januari 2017.

4.3 Variabel Penelitian dan Definisi Operasional

Variabel penelitian terdiri dari *variabel dependent* dan *variabel independent*

a. Variabel dependent

Variabel dependent atau variabel terikat dalam penelitian ini adalah tingkat keparahan luka korban kecelakaan yang terjadi di Kabupaten Sleman Provinsi DIY tahun 2016-2017 yang dikategori menjadi 3 yaitu :

1. Korban Meninggal dunia
2. Korban Luka Berat
3. Korban Luka Ringan

b. Variabel independent

Variabel independent pada penelitian ini terdiri dari 6 variabel, yaitu :

1. Waktu Kejadian

Waktu kejadian kecelakaan terbagi menjadi 2 kategori yaitu :

- a) Padat Kendaraan (antara pukul 06.00 WIB – 08.00 WIB, antara pukul 12.00 WIB – 13.30 WIB, antara pukul 16.00 WIB – 18.00 WIB)
- b) Sepi (selain waktu padat)

2. Jenis Kelamin

Jenis Kelamin korban dibagi menjadi 2 kategori yaitu :

- a) Laki-laki
- b) Perempuan

3. Usia Korban

Usia korban kecelakaan dibagi menjadi 4 kategori usia, antara lain :

- a) Anak-anak dan Remaja (1-15 tahun)
- b) Muda (16-35 tahun)
- c) Dewasa (36-55 tahun)
- d) Lanjut Usia (>55)

4. Kepemilikan SIM

Kepemilikan SIM dibagi menjadi 2 kategori yaitu :

- a) Ada
- b) Tidak Ada

5. Jenis Kecelakaan

Jenis Kecelakaan yang terjadi di Kabupaten Sleman dibagi menjadi 6 kategori yaitu :

- a) Depan – Depan
- b) Depan – Samping
- c) Depan – Belakang
- d) Samping – Samping
- e) Laka Tunggal

6. Kendaraan yang terlibat

Ada banyak jenis kendaraan yang terlibat pada kasus kecelakaan lalu lintas di Kabupaten Sleman dalam kurun waktu 2016, antara lain :

- a) Sepeda Motor (Spm)
- b) Spm - Spm
- c) Spm – Mobil
- d) Spm – Sepeda Onthel

- e) Spm – Truk
- f) Spm – Pejalan kaki
- g) Spm – Bus
- h) Spm – Traktor
- i) Mobil – Mobil
- j) Mobil – Pejalan Kaki
- k) Mobil – Sepeda Onthel
- l) Mobil – Truk
- m) Mobil – Bus
- n) Truk – Pejalan Kaki
- o) Bus – Bus
- p) Bus – Truk
- q) Bus – Pejalan Kaki

4.4. Metode Penelitian

Metode pengolahan data yang digunakan yaitu Support Vector Machine (SVM) dengan fungsi Polynomial kernel menggunakan data asli sebanyak yang telah di rebalance oleh SMOTE. Software yang digunakan dalam penelitian ini adalah Microsoft Excel 2016 dan WEKA.

4.5. Langkah Penelitian

Tahapan yang dilakukan pada penelitian ini adalah seperti pada gambar berikut :



Gambar 4.1 Alur Penelitian

BAB V ANALISIS DAN PEMBAHASAN

5.1. Analisis Deskriptif

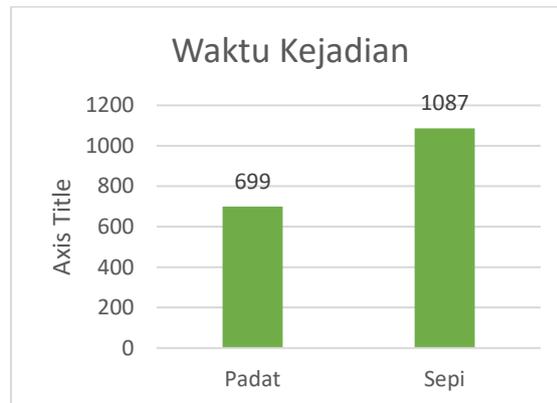
Analisis deskriptif digunakan untuk memperoleh gambaran data secara umum. Berikut adalah gambaran umum dari data sekunder kecelakaan lalu lintas di Kabupaten Sleman tahun 2016-2017.



Gambar 5.1 Jumlah kasus korban kecelakaan tahun 2016-2017
menurut tingkat keparahan luka

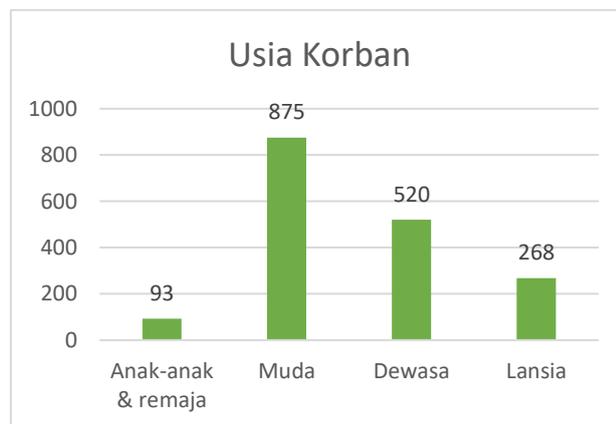
Pada **gambar 5.1** diketahui jumlah keseluruhan kasus kecelakaan lalu lintas di Kabupaten Sleman tahun 2016-2017 sebanyak 1756 dengan tingkat keparahan korban kecelakaan dengan kategori luka ringan sebanyak 1576 orang; luka berat sebanyak 8 orang dan meninggal dunia sebanyak 172 orang.

Berikut ini adalah jumlah kejadian kecelakaan lalu lintas menurut waktu kejadian :



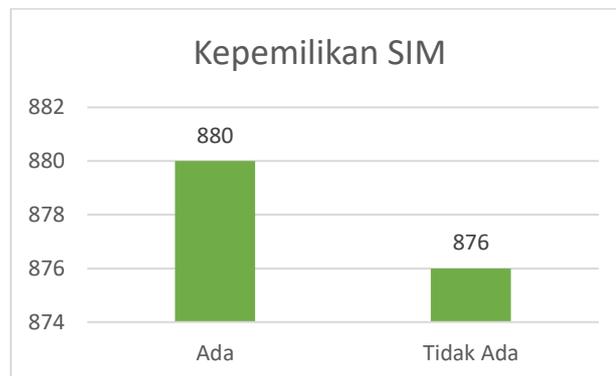
Gambar 5.2 Waktu Kejadian

Dari **Gambar 5.2** dapat dilihat bahwa kejadian kecelakaan lalu lintas yang tercatat di Sleman pada tahun 2016-2017 paling sering terjadi pada waktu sepi dibanding saat waktu lalu lintas padat. Pada saat waktu lalu lintas sepi ada 1087 kejadian kecelakaan sedangkan pada waktu lalu lintas padat yaitu sebanyak 699 kejadian.



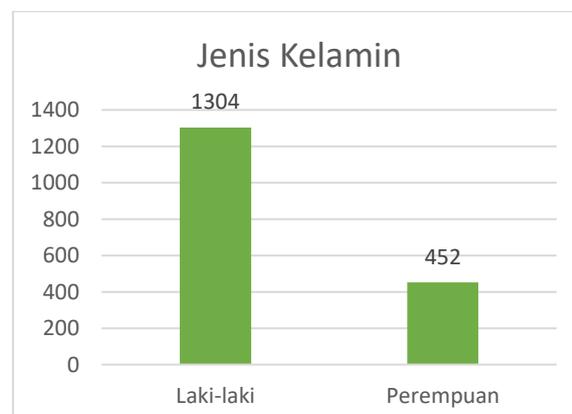
Gambar 5.3 Usia Korban

Dari **Gambar 5.3** di atas dapat dilihat jumlah korban berdasarkan usia korban yang terbagi menjadi 4 kategori yaitu Anak-anak & Remaja, Muda, Dewasa dan Lanjut Usia (lansia). Grafik menunjukkan bahwa usia korban yang paling sering terlibat kecelakaan di Sleman tahun 2016-2017 adalah pada usia muda sebagai korban kecelakaan lalu lintas yaitu sebanyak 875 korban. Kemudian urutan kedua yaitu usia dewasa sebanyak 520 korban, Lanjut Usia sebanyak 268 korban dan terakhir anak-anak & remaja sebanyak 93 korban.



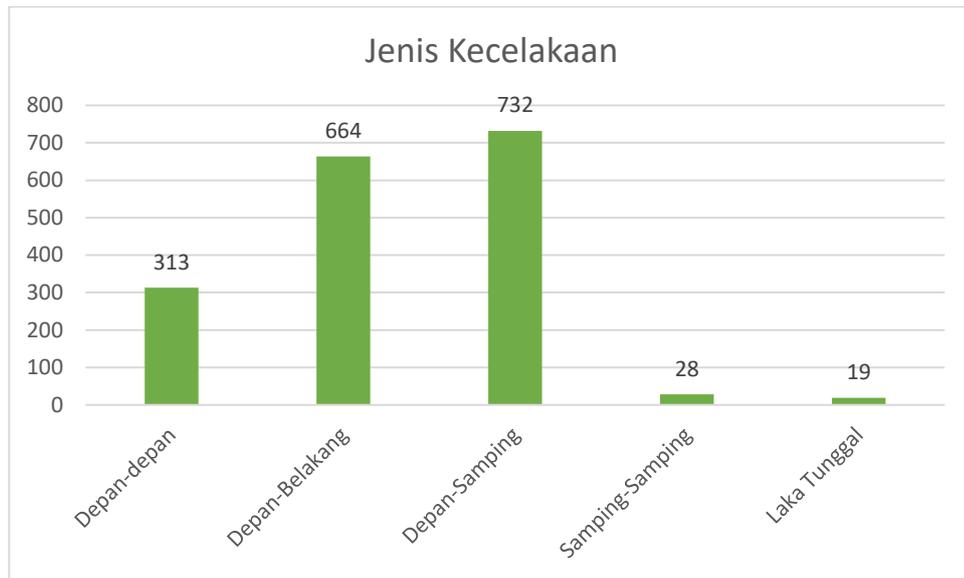
Gambar 5.4 Kepemilikan Surat Izin Mengendarai (SIM)

Dari **Gambar 5.4** dapat dilihat jumlah korban kecelakaan lalu lintas di Sleman tahun 2016-2017 menurut kepemilikan SIM. Korban kecelakaan lalu lintas dilihat dari sisi pengemudi untuk mengidentifikasi ada atau tidaknya SIM. Berdasarkan grafik di atas, diketahui ada 880 pengemudi memiliki SIM yang menjadi korban kecelakaan lalu lintas, sedangkan untuk pengemudi yang tidak memiliki SIM sebanyak 876. Jumlah ini menggambarkan bahwa pengemudi yang memiliki SIM belum menjamin keselamatan pengemudi.



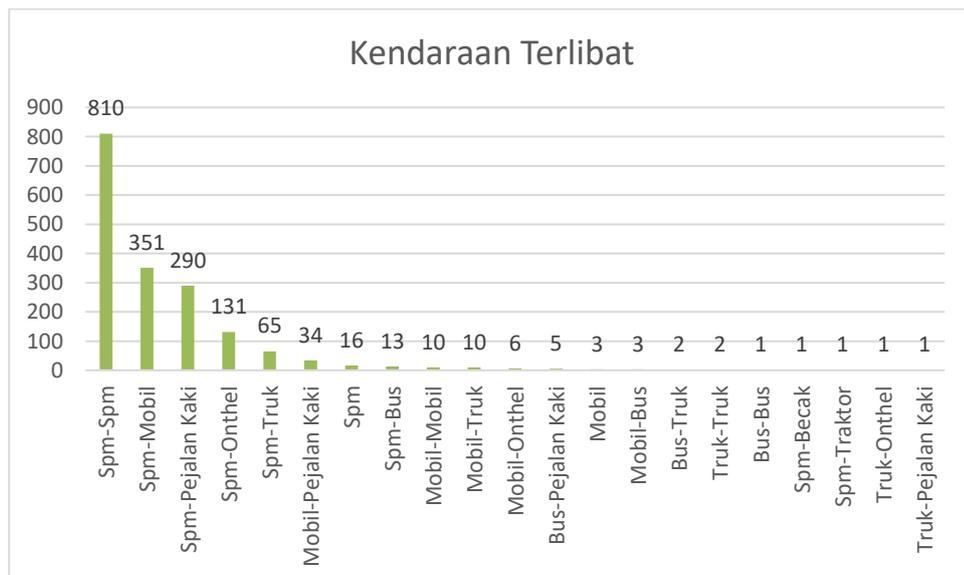
Gambar 5.5 Jenis Kelamin

Dari **gambar 5.5** dapat dilihat jumlah korban kecelakaan lalu lintas di Sleman tahun 2016-2017 menurut jenis kelamin. Pengemudi laki-laki lebih sering terlibat kecelakaan lalu lintas dibandingkan pengemudi perempuan. Tercatat ada 1304 jumlah korban kecelakaan lalu lintas berjenis kelamin laki-laki di Sleman tahun 2016-2017 sedangkan yang berjenis kelamin perempuan tercatat ada 452 korban.



Gambar 5.6 Jenis Kecelakaan

Dari **gambar 5.6** dapat dilihat jumlah kecelakaan yang terjadi menurut jenisnya. Terdapat 5 jenis kecelakaan yang dikategorikan yaitu depan-depan, depan-belakang, depan-samping, samping-samping dan laka tunggal. Jenis kecelakaan yang sering terjadi adalah depan-samping yaitu sejumlah 732 kasus kecelakaan, lalu diikuti jenis kecelakaan depan belakang sebanyak 664 kasus, depan-depan sebanyak 313 kasus, samping-samping 28 kasus dan terakhir 19 kasus kecelakaan laka tunggal.



Gambar 5.7 Kendaraan Terlibat

Dari **gambar 5.7** dapat dilihat jumlah kecelakaan yang terjadi menurut jumlah kendaraan terlibat. Terlihat bahwa kecelakaan lalu lintas yang paling sering terjadi adalah antara sepeda motor dengan sepeda motor (Spm-Spm) yaitu mencapai 810 kejadian melebihi seluruh jumlah kejadian pada jenis kendaraan terlibat yang lainnya.

Beberapa diagram batang di atas merupakan gambaran umum tiap variabel penelitian, Berikut merupakan table *crosstab* (tabel kontingensi) yang digunakan untuk mengetahui gambaran hubungan beberapa variabel dari data sekunder kecelakaan lalu lintas Kabupaten Sleman tahun 2016-2017 :

Tabel 5.1 Crosstab Jumlah Tingkat Keparahan Luka Menurut Waktu Kejadian

Luka	Waktu Kejadian		Total
	Padat	Sepi	
LR	612	964	1576
LB	0	8	8
MD	57	115	172
Total	669	1087	1756

Dari **tabel 5.1** didapatkan informasi jumlah korban tingkat keparahan luka menurut waktu kejadian. Untuk waktu kejadian padat lalu lintas sebanyak 612 korban mengalami luka ringan (LR) dan 57 korban meninggal dunia (MD). Lalu untuk kejadian sepi sebanyak 964 jiwa mengalami luka ringan, 8 orang mengalami luka berat dan 115 jiwa meninggal dunia.

Tabel 5.2 Crosstab Jumlah Tingkat Keparahan Luka Menurut Usia Korban

Luka	Usia Korban				Total
	Anak-anak & Remaja	Muda	Dewasa	Lanjut Usia	
LR	88	795	471	222	1576
LB	0	3	3	2	8
MD	5	77	46	44	172
Total	93	875	520	268	1756

Dari **tabel 5.2** didapatkan informasi jumlah korban tingkat keparahan luka menurut kategori umur. Kategori umur dibagi menjadi 4 yaitu Anak-anak dan

remaja (1 - 15 tahun) terdapat 88 korban mengalami luka ringan dan 5 orang meninggal dunia, usia muda (16 - 35 tahun) ada 795 mengalami luka ringan, 3 orang mengalami luka berat dan 77 orang meninggal dunia, usia dewasa (36 – 55 tahun) ada 471 orang mengalami luka ringan, 3 orang mengalami luka berat dan 46 orang meninggal dunia, kemudian terakhir usia lanjut usia (55 >) ada 222 orang mengalami luka ringan, 2 orang mengalami luka berat dan 77 orang meninggal dunia.

Tabel 5.3 Crosstab Jumlah Tingkat Keparahan Luka Menurut Kepemilikan SIM

Luka	Kepemilikan SIM		Total
	Ada	Tidak Ada	
LR	805	771	1567
LB	4	4	8
MD	71	101	172
Total	880	876	1756

Dari **tabel 5.3** didapatkan informasi jumlah korban tingkat keparahan luka menurut kepemilikan SIM (Surat Izin Mengendarai). Sebanyak 880 orang yang diklasifikasikan memiliki SIM, dengan 805 orang mengalami luka ringan, 4 orang mengalami luka berat dan 71 orang meninggal dunia, sedangkan ada sebanyak 876 orang yang tidak memiliki SIM dengan jumlah korban mengalami luka ringan sebanyak 771 orang, luka berat sebanyak 4 orang dan meninggal dunia sebanyak 101 orang.

Tabel 5.4 Crosstab Jumlah Tingkat Keparahan Luka Menurut Jenis Kelamin

Luka	Jenis Kelamin		Total
	Laki-laki	Perempuan	
LR	1174	402	1576
LB	7	1	8
MD	123	49	172
Total	1304	452	1756

Dari **tabel 5.4** didapatkan informasi jumlah korban tingkat keparahan luka menurut jenis kelamin laki-laki. Dari 1304 korban laki-laki, ada 1174 orang yang mengalami luka ringan, 7 orang yang mengalami luka berat dan 123 orang meninggal dunia. Kemudian dari 452 korban perempuan, ada 402 orang yang

mengalami luka ringan, 1 orang yang mengalami luka berat dan 49 orang meninggal dunia.

Tabel 5.5 Crosstab Jumlah Tingkat Keparahan Luka Menurut Jenis Kecelakaan

Luka	Jenis Kecelakaan					Total
	Depan-Belakang	Depan-Depan	Depan-Samping	Samping-Samping	Laka Tunggal	
LR	598	292	649	24	13	1576
LB	4	1	3	0	0	8
MD	62	20	80	4	6	172
Total	664	313	732	28	19	1754

Dari **tabel 5.5** didapatkan informasi jumlah korban tingkat keparahan luka menurut jenis kecelakaan. Untuk kategori kecelakaan depan-belakang sejumlah 664 kejadian dengan korban luka ringan sebanyak 598 orang, luka berat sebanyak 4 orang dan meninggal dunia 62 orang, kategori kecelakaan depan-depan sejumlah 313 kejadian dengan korban luka ringan sebanyak 292, luka berat 1 orang dan meninggal dunia 20 orang, kategori depan-samping sejumlah 732 kejadian dengan korban luka ringan sebanyak 649 orang, luka berat sebanyak 3 orang dan meninggal dunia 80. Untuk kejadian laka tunggal ada 19 kejadian dengan 13 orang luka ringan dan 6 orang meninggal, kategori samping-samping ada 28 orang dengan 24 orang luka ringan dan 4 orang meninggal.

Tabel 5.6 Crosstab Jumlah Tingkat Keparahan Luka Menurut Kendaraan Terlibat

Kendaraan Terlibat	Luka			Total
	LR	LB	MD	
Spm-Spm	775	3	32	810
Spm-Mobil	319	0	32	351
Spm-Pejalan Kaki	241	1	48	290
Spm-Onthel	119	2	10	131
Spm-Truk	46	0	19	65
Mobil-Pejalan Kaki	20	1	13	34
Spm	10	0	6	16
Spm-Bus	10	0	3	13
Mobil-Mobil	9	0	1	10
Mobil-Truk	6	1	3	10
Mobil-Onthel	5	0	1	6
Bus-Pejalan Kaki	4	0	1	5
Mobil	3	0	0	3
Mobil-Bus	3	0	0	3
Bus-Truk	2	0	0	2
Truk-Truk	1	0	1	2
Bus-Bus	1	0	0	1
Spm-Becak	0	0	1	1
Spm-Traktor	1	0	0	1
Truk-Onthel	0	0	1	1
Truk-Pejalan Kaki	1	0	0	1
Total	1576	8	172	1756

5.2. Analisis Data dengan Metode SVM

Analisis menggunakan data kecelakaan lalu lintas di Kabupaten Sleman tahun 2016-2017 dengan jumlah data sebanyak 1756 data korban . Pada penelitian ini menggunakan SVM yang *non-linear* dengan fungsi *Kernel Polynomial*.

5.2.1 Klasifikasi menggunakan SVM *polynomial Kernel*

SVM dengan fungsi SVM *polynomial Kernel* menggunakan nilai cost yaitu 1. Dimana nilai cost tersebut diterapkan pada 1756 data asli sebelum digunakan metode SMOTE. Selanjutnya nilai cost tersebut diterapkan untuk mendapatkan confusion matrix. Berikut hasil output confusion matrix dari klasifikasi dataset :

Tabel 5.7 Matrik Konfusi

Luka	Luka Ringan	Meninggal Dunia	Luka Berat	Total
Luka Ringan	1574	2	0	1576
Meninggal Dunia	167	5	0	177
Luka Berat	7	1	0	8
Total	1748	8	0	1756

Dengan :

$$\text{Akurasi} : \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{1574+5+0}{1756} = \frac{1579}{1756} = 89.920 \%$$

$$\text{Error} : \frac{\sum \text{Prediksi salah}}{\sum (\text{benar} + \text{salah})} = \frac{2+167+8}{1748} = \frac{177}{1748} = 10.079 \%$$

$$\text{Presisi LR} : \frac{\text{Prediksi benar LR}}{\text{Prediksi benar dan salah LR}} = \frac{1574}{1574+2} = \frac{1574}{1576} = 99.873\%$$

$$\text{Presisi MD} : \frac{\text{Prediksi benar MD}}{\text{Prediksi benar dan salah MD}} = \frac{5}{167+5+0} = 2.824\%$$

$$\text{Presisi LB} : \frac{\text{Prediksi benar LB}}{\text{Prediksi benar dan salah LB}} = \frac{0}{7+1+0} = 0$$

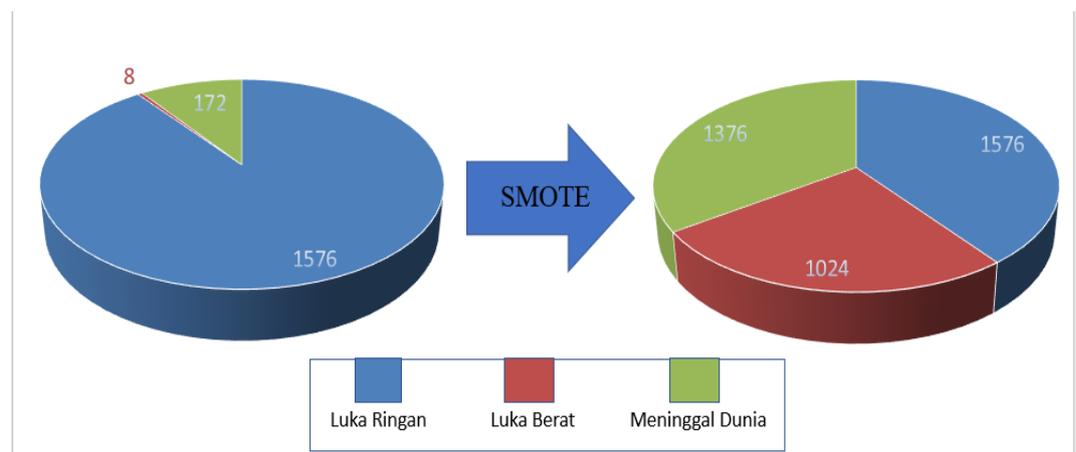
Pada tabel 5.7 di atas didapatkan dari 1756 data kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 di klasifikasi dengan metode *SVM polynomial Kernel*. ada sebanyak 1579 data yang berhasil diklasifikasi dengan benar sedangkan sebanyak 177 data tidak diklasifikasi dengan benar. Berdasarkan perhitungan bahwa persentase ketepatan klasifikasi atau nilai keakuratan dalam menganalisis data keparahan luka korban kecelakaan lalu lintas Kabupaten Sleman adalah sebesar 89.920 % dengan tingkat kesalahan klasifikasi sebesar 10.079 %. Kemudian untuk tingkat presisi nya masing-masing adalah untuk kategori luka ringan (LR) sebesar 99.873 % , kategori luka berat (LB) sebesar 0 % dan kategori meninggal dunia (MD) sebesar 2.029 %. Hal ini dikarenakan adanya ketidakseimbangan kelas pada data sehingga klasifikasi lebih condong ke kelas mayoritas yaitu kelas Luka Ringan yang memang tercatat dengan jumlah yang tinggi daripada kelas luka berat dan meninggal dunia. Hasil klasifikasi pada tabel 5.6 tentunya tidak bisa digunakan untuk pihak yang bersangkutan sebagai informasi

dalam mengetahui peluang kecelakaan di Sleman Yogyakarta karena tingkat keakuratannya yang tidak real.

Untuk mengatasi hal tersebut dilakukan suatu pendekatan *rebalancing* pada dataset dengan menggunakan teknik sampling *SMOTE* yang kemudian digabungkan dengan klasifikasi *SVM polynomial Kernel* dan *Radial Basis Function (RBF)*

5.2.2 Klasifikasi menggunakan *SVM polynomial Kernel* dengan metode *SMOTE (Synthetic Minority Over-sampling Technique)*

Diketahui bahwa data asli yaitu sejumlah 1756 data kejadian lalu lintas yang kemudian dilakukan oversampling menggunakan teknik SMOTE, sehingga data mengalami penambahan hingga berjumlah 3976 data kejadian kecelakaan lalu lintas.



Gambar 5.8 Ilustrasi perubahan jumlah data setelah oversampling menggunakan SMOTE

SVM dengan fungsi *Polynomial SMOTE* diterapkan pada 3976 data setelah dilakukan Teknik sampling *SMOTE* untuk mengatasi *Imbalance dataset*. Pada klasifikasi ini dilakukan percobaan menggunakan fungsi kernel Polynomial dengan menggunakan dua nilai d pada fungsi Kernel Polynomial yaitu $d = 1$ dan $d = 2$ dengan nilai parameter $C = 1$, $C = 5$, $C = 10$, $C = 50$, $C = 100$ dan akan dibandingkan ketepatan klasifikasi untuk masing-masing parameter fungsi kernel polynomial. Berikut adalah tabel hasil akurasi untuk prediksi kelas pada data training.

Tabel 5.8 Hasil Akurasi Klasifikasi SVM Kernel *Polynomial* pada Data *Training*

C	Polynomial	
	d = 1	d = 2
1	77.238 %	83.652 %
5	77.288 %	83.803 %
10	77.288 %	83.828 %
50	77.264 %	83.828 %
100	77.264 %	83.828 %

Dari tabel 5.7 didapatkan hasil akurasi terbaik dalam pengklasifikasian adalah sebesar 83.828 % untuk parameter $d = 2$ dan $C = 10$ Selanjutnya nilai parameter dan cost tersebut diterapkan untuk mendapatkan *confusion matrix*. Berikut adalah tabel hasil akurasi untuk prediksi kelas pada data testing.

Tabel 5.9 Hasil Akurasi Klasifikasi SVM Kernel *Polynomial* pada Data *Testing*

C	Polynomial	
	d = 1	d = 2
1	76.101 %	82.264 %
5	76.101 %	82.138 %
10	76.101 %	82.264 %
50	76.101 %	82.264 %
100	76.101 %	82.264 %

Dari tabel 5.9 didapatkan hasil akurasi terbaik dalam pengklasifikasian adalah sebesar 82.264 % untuk parameter $d = 2$ dan $C = 10$ Selanjutnya nilai parameter dan cost tersebut diterapkan untuk mendapatkan *confusion matrix*. Berikut hasil output *confusion matrix* dari klasifikasi *dataset* :

Tabel 5.10 Matrik Konfusi SVM Polynomial Dengan SMOTE Pada Data Training

Luka	Luka Ringan	Luka Berat	Meninggal Dunia	Total
Luka Ringan	1220	313	43	1576
Meninggal Dunia	242	1096	38	1376
Luka Berat	6	1	1017	1024
Total	1468	1410	1098	3976

Dengan :

$$\text{Akurasi} : \frac{\sum \text{Prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{1220 + 1096 + 1017}{3976} = \frac{3333}{3976} = 83.828 \%$$

$$\text{Error} : \frac{\sum \text{Prediksi salah}}{\sum (\text{benar} + \text{salah})} = \frac{356 + 280 + 7}{3976} = \frac{643}{3976} = 16.172 \%$$

$$\text{Presisi LR} : \frac{\text{Prediksi benar LR}}{\text{Prediksi benar dan salah LR}} = \frac{1220}{1220 + 313 + 43} = \frac{1220}{1576} = 77.41 \%$$

$$\text{Presisi LB} : \frac{\text{Prediksi benar LB}}{\text{Prediksi benar dan salah LB}} = \frac{1017}{6 + 1 + 1017} = \frac{1017}{1024} = 99.31 \%$$

$$\text{Presisi MD} : \frac{\text{Prediksi benar MD}}{\text{Prediksi benar dan salah MD}} = \frac{1096}{242 + 1096 + 38} = \frac{1096}{1376} = 79.65 \%$$

Pada tabel 5.10 di atas didapatkan dari 3976 data kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 ada sebanyak 3333 data yang berhasil diklasifikasi dengan benar sedangkan sebanyak 643 data diklasifikasi berbeda dari kelas asli. Berdasarkan perhitungan bahwa persentase ketepatan klasifikasi atau nilai keakuratan dalam menganalisis data keparahan luka korban kecelakaan lalu lintas Kabupaten Sleman adalah sebesar 83.828 % dengan tingkat kesalahan klasifikasi sebesar 16.172 %. Kemudian untuk tingkat presisi nya masing-masing adalah untuk kategori luka ringan (LR) sebesar 77.41 %, kategori luka berat (LB) sebesar 99.31 % dan kategori meninggal dunia (MD) sebesar 79.65 %.

Tabel 5.11 Matrik konfusi SVM Polynomial dengan SMOTE pada data testing

Luka	Luka Ringan	Luka Berat	Meninggal Dunia	Total
Luka Ringan	233	77	6	316
Meninggal Dunia	43	218	12	273
Luka Berat	2	1	203	206
Total	278	296	221	795

Dengan :

$$\text{Akurasi} : \frac{\sum \text{prediksi benar}}{\sum (\text{benar} + \text{salah})} = \frac{233 + 218 + 203}{795} = \frac{654}{795} = 82.264 \%$$

$$\text{Error} : \frac{\sum \text{Prediksi salah}}{\sum (\text{benar} + \text{salah})} = \frac{83 + 55 + 3}{795} = \frac{390}{795} = 17.735 \%$$

$$\text{Presisi LR} : \frac{\text{Prediksi benar LR}}{\text{Prediksi benar dan salah LR}} = \frac{233}{233 + 77 + 6} = \frac{233}{316} = 73.73 \%$$

$$\text{Presisi LB} : \frac{\text{Prediksi benar LB}}{\text{Prediksi benar dan salah LB}} = \frac{203}{2 + 1 + 203} = \frac{204}{206} = 98.54 \%$$

$$\text{Presisi MD} : \frac{\text{Prediksi benar MD}}{\text{Prediksi benar dan salah MD}} = \frac{218}{43 + 218 + 12} = \frac{218}{273} = 79.85 \%$$

Pada tabel 5.11 di atas didapatkan dari data testing data kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 yaitu 20 % dari 3976 kejadian didapatkan 795 data, ada sebanyak 654 data yang berhasil diklasifikasi dengan benar sedangkan sebanyak 141 data tidak diklasifikasi dengan benar. Berdasarkan perhitungan bahwa persentase ketepatan klasifikasi atau nilai keakuratan dalam menganalisis data keparahan luka korban kecelakaan lalu lintas Kabupaten Sleman adalah sebesar 82.264 % dengan tingkat kesalahan klasifikasi sebesar 17.735 %. Kemudian untuk tingkat presisi nya masing-masing adalah untuk kategori luka ringan (LR) sebesar 73.73 %, kategori luka berat (LB) sebesar 98.54 % dan kategori meninggal dunia (MD) sebesar 79.85 %.

5.3 Validasi Metode

Dari metode SVM dengan fungsi *Polynomial Kernel* yang telah di-rebalance menggunakan SMOTE, didapatkan hasil tingkat akurasi untuk dilakukan validasi metode dengan fungsi *Kernel*. Penentuan fungsi *Kernel* yang baik untuk digunakan dilihat dari nilai ketepatan klasifikasi yang tinggi.

Tabel 5.12 Perbandingan Tingkat Akurasi

Fungsi Kernel	Akurasi Klasifikasi
<i>Polynomial SMOTE</i>	82.264 %
<i>RBF SMOTE</i>	74.2138 %
<i>Sigmoid SMOTE</i>	74.0881 %

Tabel 5.13 Perbandingan Tingkat Presisi

Kelas	Perbandingan Presisi Klasifikasi		
	Polynomial SMOTE	RBF SMOTE	Sigmoid SMOTE
LR	73.73 %	71.5 %	71.2 %
LB	98.54 %	99.02 %	99.02 %
MD	79.85 %	58.6 %	58.6 %

Berdasarkan tabel 5.13 dapat dilihat bahwa dari kedua fungsi *Kernel* yang digunakan terdapat fungsi *Kernel Polynomial* memiliki tingkat akurasi yang lebih tinggi dibandingkan fungsi RBF dan Sigmoid. Apabila dilihat dari tingkat presisi ketiga fungsi kernel tersebut, didapatkan persamaan tingkat presisi pada kelas luka berat (LB) yaitu sebesar 99.02 % pada kernel RBF dan Sigmoid sedangkan pada kernel Polynomial sebesar 98.54. Pada tingkat presisi kelas kelas luka ringan (LR) yaitu sebesar 73.73 % pada fungsi Polynomial kernel, 71.5 % pada fungsi Radial Basis Function (RBF) dan 71.2 % pada fungsi Sigmoid. Lalu untuk kelas meninggal dunia (MD) yaitu sebesar 79.85 % pada fungsi *polynomial kernel*, 58.6 % pada fungsi *radial basis function (RBF)* dan fungsi *Sigmoid*.

BAB VI

PENUTUP

6.1. Kesimpulan

Berdasarkan pembahasan diperoleh kesimpulan bahwa dari hasil analisis yang dilakukan, terdapat beberapa informasi yaitu :

1. Gambaran umum pada data kecelakaan di Kabupaten Sleman tahun 2016-2017 yaitu diketahui jumlah keseluruhan kasus kecelakaan lalu lintas di Kabupaten Sleman sebanyak 1756 dengan tingkat keparahan korban kecelakaan dengan kategori luka ringan sebanyak 1576 orang; luka berat sebanyak 8 orang dan meninggal dunia sebanyak 172 orang.
2. Hasil klasifikasi tingkat keparahan luka korban kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 menggunakan metode SVM (Support Vector Machine) *Polynomial Kernel SMOTE* adalah sebagai berikut :

- a. Data Training

Dari 3976 data SMOTE kecelakaan lalu lintas di Kabupaten Sleman, Yogyakarta tahun 2016-2017 ada sebanyak 3071 data yang berhasil diklasifikasi dengan benar sedangkan sebanyak 905 data diklasifikasi berbeda dengan data asli. Berdasarkan perhitungan bahwa persentase ketepatan klasifikasi atau nilai keakuratan dalam menganalisis data keparahan luka korban kecelakaan lalu lintas Kabupaten Sleman adalah sebesar 83.828 % dengan tingkat kesalahan klasifikasi sebesar 16.172 %. Kemudian untuk tingkat presisi nya masing-masing adalah untuk kategori luka ringan (LR) sebesar 77.41 %, kategori luka berat (LB) sebesar 99.31 % dan kategori meninggal dunia (MD) sebesar 79.65 %.

- b. Data Testing

Dari 795 data testing SMOTE, ada sebanyak 654 data yang berhasil diklasifikasi dengan benar sedangkan sebanyak 141 data diklasifikasi berbeda dari kelas asli. Berdasarkan perhitungan

bahwa persentase ketepatan klasifikasi atau nilai keakuratan dalam menganalisis data keparahan luka korban kecelakaan lalu lintas Kabupaten Sleman adalah sebesar 82.264 % dengan tingkat kesalahan klasifikasi sebesar 17.735 %. Kemudian untuk tingkat presisi nya masing-masing adalah untuk kategori luka ringan (LR) sebesar 73.73 %, kategori luka berat (LB) sebesar 98.54 % dan kategori meninggal dunia (MD) sebesar 79.85 %.

6.2 Saran

Berdasarkan kesimpulan diatas, maka dapat dikemukakan beberapa saran, yaitu :

1. Diharapkan untuk pihak kepolisian Sleman agar mengoptimalkan kegiatan rekapitulasi data kecelakaan lalu lintas di Sleman agar tidak terdapat data missing dalam penelitian selanjutnya.
2. Untuk penelitian selanjutnya agar penelitian serupa diperlukan penambahan variabel-variabel yang berpengaruh pada data kecelakaan lalu lintas di Kabupaten Sleman guna mendapatkan pola dan klasifikasi yang lebih akurat pada data kecelakaan tahun selanjutnya.
3. Penelitian selanjutnya diharapkan menggunakan metode oversampling yang lebih baik untuk mengatasi *Imbalance data*.

DAFTAR PUSTAKA

- Adisasmita, S.A. 2011. Jaringan Transportasi Teori dan Analisis. Yogyakarta: Graha Ilmu,
- Anonim. 2017. Vektor. <http://staffnew.uny.ac.id/upload/131808335/pendidikan/VEKTOR.pdf> (di akses tanggal 23 Oktober 2017 pukul 20:30).
- Assegaf, J.S., 2016, Sepanjang 2015, 27.000 Orang Meninggal di Jalan <http://www.solopos.com/2016/04/10/angka-kecelakaan-sepanjang-2015-27-000-orang-meninggal-di-jalan-708925> diakses tanggal 1 Oktober 2017 pukul 22.49 WIB.
- Athoillah, M., Irawan, M.I dan Imah, E.M.,. 2015. Support Vector Machine Untuk Image Retrieval. Prosiding Seminar Nasional Matematika dan Pendidikan Matematika 2015 ISBN No. 978-979-028-728-0.
- Azizirrahman, M., Normelani, E dan Arisanty, D. 2015. Faktor Penyebab Terjadinya Kecelakaan Lalu Lintas pada Daerah Rawan Kecelakaan di Kecamatan Banjarmasin Tengah Kota Banjarmasin. JPG (Jurnal Pendidikan Geografi) Volume 2, No 3, Mei 2015: 20-37.
- Bustami. 2014. Penerapan Algoritma Naïve Bayes untuk mengklasifikasi data nasabah asuransi. Jurnal Informatika. Vol. 8, No. 1, Januari 2014.
- DPR, Undang - Undang RI Nomor 22 Tahun 2009 Tentang Lalu Lintas dan Angkutan Jalan & Peraturan Pemerintah RI Nomor 55 Tahun 2012 Tentang Kendaraan Beserta Penjelasannya, Surabaya: Kesindo Utama, 2012.
- Durgesh. 2009. Data Classification Using Support Vector Machine. Journal of Theoretical and Applied Information Technology.
- Fernandy Marbun, Z.K. 2009. Analisis Dan Implementasi Ensemble Of Support Vector Machine Pada Kasus Churn Prediction Pelanggan Telekomunikasi. Tugas Akhir – 2009.
- Haryati, S., Sudarsono, A. dan Suryana, E. 2015. Implementasi Data Mining untuk Memprediksi Masa Studi Mahasiswa menggunakan Algoritma C4.5 (Studi Kasus : Universitas Dehasen Bengkulu). Jurnal Media Infotama. Vol. 11 (2). 130-138.
- Hossin, M. 2015. A Review On Evaluation Metrics For Data Classification Evaluations. International Journal of Data Mining & Knowledge Management Process (IJDMP) Vol.5, No.2, March 2015.

- Hsu, C.W., Chang, C.C. dan Lin, C.J. 2003. *A Practical Guide to Support vector Classification*. Taiwan: Department of Computer Science National Taiwan University.
- Ispriyanti, D dan Hoyyi, A. 2016. Analisis Klasifikasi Masa Studi Mahasiswa Prodi Statistika Undip Dengan Metode Support Vector Machine (Svm) Dan Id3 (Iterative Dichotomiser 3). *Media Statistika* 9(1) 2016: 15-29.
- Larose D, T. 2006. *Data Mining Methods and Models*, Jhon Wiley & Sons. Inc. Hoboken New Jersey
- Nugroho, A.S., Witarto, A.B, dan Handoko, D. 2003. Support Vector Machine – Teori dan Aplikasinya dalam Bioinformatika. *Jurnal. IlmuKomputer.com*.
- Octaviani, P.A., Wilandari, Y. dan Ispriyanti, D, 2014. Penerapan Metode Klasifikasi Support Vector Machine (Svm) Pada Data Akreditasi Sekolah Dasar (Sd) Di Kabupaten Magelang. *Jurnal Gaussian*, Volume 3, Nomor 4, Tahun 2014, Halaman 811 – 820.
- Ogor, Emmanuel N. (2007). *Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques*. Fourth Congress of Electronics, Robotics and Automotive Mechanics.
- Pemerintah Republik Indonesia, Peraturan Pemerintah Republik Indonesia Nomor 43 Tahun 1993 Tentang Prasarana Dan Lalu Lintas Jalan, Jakarta: Departemen Perhubungan, 1993.
- Putri, Y.D. 2017. *Klasifikasi Status Kelulusan Mahasiswa Menggunakan Metode Support Vector Machine*. Skripsi. Program Studi Statistika. Universitas Islam Indonesia.
- Riquelme, J.C., Ruiz, R., Rodriguez, D. dan Moreno, J. 2008. Finding Defective Modules from Highly Unbalanced Datasets. *Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos*, Vol. 2, No. 1, 2008.
- Susilowati, E., Sabariah, M.K. dan Gozali A. Asusilowati . 2015. Implementasi Metode Support Vector Machine untuk melakukan Klasifikasi Kemacetan Lalu Lintas pad Twitter. *Jurnal. Universitas Telkom*.
- Turban, E., Aronson, J.E. dan Liang, T.P. 2005. *Decision Support Systems and Intelligent Systems* 7th Ed. New Jersey : Pearson Education.

LAMPIRAN

Lampiran 1 Sampel Data Kecelakaan Lalu Lintas Sleman, Yogyakarta Tahun 2016-2017.

No.	Waktu Kejadian	Kendaraan Terlibat	Umur	SIM	JK	Jenis Kecelakaan	Luka
1	Sepi	Spm-Pejalan Kaki	Muda	Ada	Laki-laki	Depan-Samping	LR
2	Padat	Spm-Onthel	Lansia	Tidak Ada	Laki-laki	Depan-Belakang	LR
3	Sepi	Spm-Pejalan Kaki	Muda	Ada	Perempuan	Depan-Samping	LR
4	Sepi	Spm-Spm	Muda	Tidak Ada	Perempuan	Depan-Samping	LR
5	Sepi	Spm-Spm	Dewasa	Tidak Ada	Laki-laki	Depan-Samping	LR
6	Padat	Spm-Mobil	Dewasa	Ada	Laki-laki	Depan-Samping	LR
7	Sepi	Spm-Spm	Lansia	Ada	Laki-laki	Depan-Samping	LR
8	Sepi	Spm-Pejalan Kaki	Muda	Tidak Ada	Laki-laki	Depan-Samping	LR
9	Padat	Spm-Spm	Muda	Tidak Ada	Laki-laki	Depan-Samping	LR
10	Sepi	Spm-Spm	Dewasa	Ada	Perempuan	Depan-Belakang	LR
...	Sepi	Spm-Spm-Truk	Dewasa	Tidak Ada	Perempuan	Depan-Depan	MD
...	Sepi	Mobil-Pejalan kaki	Muda	Ada	Laki-laki	Depan-Samping	LB
...	Sepi	Spm-Onthel	Lansia	Tidak Ada	Laki-laki	Depan-Belakang	LB
...	Sepi	Spm-Spm	Dewasa	Tidak Ada	Laki-laki	Depan-Depan	LB
...	Sepi	Spm-Onthel	Muda	Ada	Laki-laki	Depan-Belakang	MD
...	Sepi	Spm-Mobil	Muda	Ada	Laki-laki	Depan-Samping	LR
1754	Padat	Spm-Spm	Dewasa	Tidak Ada	Laki-laki	Depan-Samping	LR
1755	Sepi	Spm-Mobil	Muda	Tidak Ada	Laki-laki	Depan-Belakang	LR
1756	Sepi	Spm-Truk	Muda	Tidak Ada	Laki-laki	Depan-Samping	MD

Lampiran 2 Output Klasifikasi SVM menggunakan WEKA

Klasifikasi SVM Polynomial dari data training

```

=== Summary ===

Correctly Classified Instances      1579           89.9203 %
Incorrectly Classified Instances     177           10.0797 %
Kappa statistic                     0.0504
Mean absolute error                  0.2456
Root mean squared error              0.3122
Relative absolute error              198.4005 %
Root relative squared error          125.7692 %
Total Number of Instances           1756

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,999   0,967   0,900     0,999   0,947     0,144   0,516    0,900    LR
                0,029   0,002   0,625     0,029   0,056     0,120   0,514    0,113    MD
                0,000   0,000   0,000     0,000   0,000     0,000   0,500    0,005    LB
Weighted Avg.   0,899   0,868   0,869     0,899   0,855     0,141   0,516    0,819

=== Confusion Matrix ===

  a  b  c  <-- classified as
1574 2  0 |  a = LR
 167 5  0 |  b = MD
   7 1  0 |  c = LB

```

Klasifikasi SVM Polynomial SMOTE dari data training (d=2, C=10)

```

Correctly Classified Instances      3333           83.828 %
Incorrectly Classified Instances     643           16.172 %
Kappa statistic                     0.7549
Mean absolute error                  0.2596
Root mean squared error              0.3338
Relative absolute error              59.2783 %
Root relative squared error          71.3361 %
Total Number of Instances           3976

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,774   0,103   0,831     0,774   0,802     0,680   0,847    0,740    LR
                0,797   0,121   0,777     0,797   0,787     0,672   0,867    0,708    MD
                0,993   0,027   0,926     0,993   0,959     0,945   0,985    0,924    LB
Weighted Avg.   0,838   0,090   0,837     0,838   0,837     0,745   0,889    0,776

=== Confusion Matrix ===

  a  b  c  <-- classified as
1220 313 43 |  a = LR
 242 1096 38 |  b = MD
   6  1 1017 |  c = LB

```

Klasifikasi SVM Polynomial SMOTE dari data testing (d=2, C=10)

```

Correctly Classified Instances      654          82.2642 %
Incorrectly Classified Instances    141          17.7358 %
Kappa statistic                    0.7317
Mean absolute error                 0.2647
Root mean squared error             0.3414
Relative absolute error             60.4489 %
Root relative squared error         72.9574 %
Total Number of Instances          795

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,737   0,094   0,838     0,737   0,785     0,660   0,833    0,729    LR
          0,799   0,149   0,736     0,799   0,766     0,638   0,848    0,674    MD
          0,985   0,031   0,919     0,985   0,951     0,934   0,980    0,912    LB
Weighted Avg.  0,823   0,097   0,824     0,823   0,821     0,723   0,876    0,757

=== Confusion Matrix ===

  a  b  c  <-- classified as
233 77  6 |  a = LR
 43 218 12 |  b = MD
  2  1 203 |  c = LB

```