

Penggunaan Metode Support Vector Machine (SVM) dalam Mengidentifikasi Tingkat Keparahan Pada Kecelakaan Lalu Lintas

FATIHATUR RAHMI¹, FERRA YANUAR², YUDIANTRI ASDI³

^{1,2,3}Departemen Matematika dan Sains Data Fakultas MIPA Universitas Andalas, Indonesia
e-mail: ferrayanuar@sci.unand.ac.id

ABSTRAK

Kendaraan sudah menjadi kebutuhan pokok yang digunakan semua orang untuk berpindah dari satu tempat ke tempat lain dengan cepat. Namun, bertambahnya jumlah kendaraan juga menimbulkan dampak negatif, salah satunya adalah kecelakaan. Berdasarkan data yang diperoleh dari website Badan Pusat Statistik (BPS) Sumatera Barat pada tahun 2018-2021, kasus kecelakaan terbanyak terjadi di Kota Padang yaitu sebesar 22,79% dari semua kasus kecelakaan yang terjadi di provinsi Sumatera Barat. Tingkat keparahan yang dialami korban pasca kecelakaan dikelompokkan kedalam 3 kategori yaitu korban yang mengalami luka ringan, luka berat dan meninggal dunia. Metode klasifikasi dapat digunakan untuk mengklasifikasi tingkat keparahan kecelakaan berdasarkan faktor-faktor yang mempengaruhi terjadinya kecelakaan. Salah satu metode yang dapat digunakan adalah metode Support Vector Machine (SVM). SVM adalah suatu learning machine yang digunakan untuk mengklasifikasi data secara statistika dalam ruang fitur berdimensi tinggi dan solusi yang dihasilkan dari klasifikasi menggunakan SVM bersifat sama untuk setiap percobaan yang dilakukan. Pada penelitian ini akan digunakan klasifikasi dengan SVM multiclass dengan metode one against one (satu lawan satu) dengan dua fungsi kernel yang selanjutnya akan dilakukan perbandingan kualitas model berdasarkan akurasi, nilai APER dan F1-score. Data yang digunakan pada penelitian ini adalah data kecelakaan yang dialami pengendara sepeda motor di kota Padang pada bulan Januari-Maret 2022. Hasil penelitian menunjukkan bahwa penggunaan kernel RBF lebih baik dibanding kernel linear dengan tingkat akurasi sebesar 94,62% dengan nilai APER sebesar 5,38% dan diperoleh F1-score untuk kategori luka ringan sebesar 97,07%, luka berat sebesar 59,90% dan meninggal dunia sebesar 80%.

Kata Kunci: Support Vector Machine, One Against One, Kernel Linear, Kernel RBF.

ABSTRACT

Transportation has become a basic necessity that everyone uses to move from one place to another quickly. However, the increasing number of transportation also has negative impacts, one of them was a traffic accident. According to BPS, the highest number of accidents occurred in Padang city, which was around 22.79% of the total cases that occurred in West Sumatra. The classification method can be used to classify the severity of accidents based on the factors that influence the occurrence of accidents. One method that can be used is the Support Vector Machine (SVM). SVM is a learning machine that is used to classify data statistically in a high-dimensional feature space and the solution resulting from classification using SVM is the same for every experiment carried out. In this research, multiclass SVM classification will be used with the one-against-one method with two kernel functions, then the model quality will be calculated based on accuracy, APER value and F1 score. The data used in this research is traffic accidents by motorcyclists in Padang City in January-March 2022. The results of the research show that the RBF kernel is better than the linear kernel with an accuracy level is 94.62%, an APER value is 5.38% and a F1-score for the minor injuries category is 97.07%, while serious injuries and deaths are 59.90% and 80%.

Keywords: Support Vector Machine, One Against One, Kernel Linear, Kernel RBF.

1. PENDAHULUAN

Kendaraan bermotor adalah suatu alat transportasi yang digerakkan peralatan teknik berupa mesin sehingga dapat mempermudah aktivitas manusia untuk berpindah dari suatu tempat ke tempat lain dengan mudah, cepat, dan tidak membutuhkan banyak tenaga. Semakin bertambahnya jumlah kendaraan bermotor yang digunakan oleh masyarakat setiap tahunnya tentu menimbulkan masalah baru dalam kehidupan manusia. Kecelakaan merupakan salah satu contoh masalah yang sering terjadi sebagai dampak buruk penggunaan kendaraan bermotor. Menurut data yang diperoleh di website Badan Pusat Statistik (BPS) Sumatera Barat pada tahun 2018-2021, kasus kecelakaan terbanyak terjadi di Kota Padang yaitu 22,79% dari total kasus kecelakaan yang terjadi di provinsi Sumatera Barat. Terdapat beberapa faktor yang menyebabkan terjadinya kecelakaan lalu lintas yaitu faktor kesalahan manusia seperti: batas muatan dan batas kecepatan; faktor pengemudi seperti mabuk, mengantuk, dan lalai; faktor jalan seperti kerusakan jalan; faktor alam seperti cuaca; dan faktor kendaraan (Enggarsari & Sa'diyah, 2017). Tingkat keparahan korban kecelakaan biasa dikelompokkan ke dalam tiga kategori, yaitu korban yang mengalami luka ringan, luka berat, dan meninggal dunia. Metode klasifikasi dapat digunakan untuk mengetahui tingkat keparahan yang alami korban berdasarkan faktor-faktor yang mempengaruhi terjadinya suatu kecelakaan. Salah satu metode klasifikasi yang dapat digunakan adalah *Support Vector Machine (SVM)*. SVM adalah suatu *learning machine* yang digunakan untuk mengklasifikasi data secara statistika dalam ruang fitur berdimensi tinggi (Cristianini & Taylor, 2014). SVM bekerja dengan mentransformasikan data pada *input space* ke ruang yang berdimensi tinggi lalu melakukan optimasi di ruang vektor baru dengan strategi *Structural Risk Minimization (SRM)* atau meminimalkan resiko struktural dimana hal ini adalah ciri-ciri SVM yang merupakan keunggulannya dibanding metode lainnya (Wibawa, et al., 2018). Selain itu solusi yang dihasilkan dari klasifikasi SVM bersifat sama untuk setiap percobaan yang dilakukan (Kusumahadi, et al., 2019). SVM bekerja dengan cara menemukan *hyperplane* (pemisah) dengan membagi data dalam 2 bagian yaitu *data training* dan *data testing* (Hastie, et al., 2008). Hyperplane dapat dirumuskan sebagai berikut,

$$f(x) = \langle w, x \rangle + b \quad \dots (1)$$

dimana w dan b masing-masing adalah vektor bobot dan bias. Kasus *non separable* dapat diselesaikan dengan menggunakan metode *soft margin classifier* dengan menambahkan variabel *slack* ϵ_i ($\epsilon_i \geq 0$) ke dalam fungsi tujuan dan kendalanya. (Scholkhopf & Smola, 2002).

$$\text{Fungsi tujuan :} \quad \min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \epsilon_i \quad \dots (2)$$

$$\text{Fungsi kendala :} \quad y_i \langle w, x_i \rangle + b \geq 1 - \epsilon_i, i = 1, 2, \dots, n$$

Menggunakan bentuk primal problem, persamaan lagrange multiplier ditulis sebagai berikut (Fide, et al., 2021),

$$L_p(w, b, \alpha_i) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \epsilon_i - \sum_{i=1}^n \alpha_i ((y_i \langle w, x_i \rangle + b) - 1 + \epsilon_i) - \sum_{i=1}^n \mu_i \epsilon_i \quad \dots (3)$$

dimana α_i dan μ_i adalah *lagrange multiplier* (pengali *lagrange*) dan ϵ_i adalah variabel *slack*. Persamaan (3) akan mencapai solusi optimal ketika memenuhi kondisi dimana persamaan (3) dapat memaksimalkan nilai α_i dan meminimalkan w dan b , sehingga persamaan (3) dapat diselesaikan jika memenuhi *Karush-Kuhn-Tucker (KKT) condition*. Sehingga berdasarkan KKT, persamaan *lagrange* dengan bentuk *dual problem* sebagai berikut (Cristianini & Taylor, 2014),

$$\text{Fungsi tujuan :} \quad L_d = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad \dots (4)$$

$$\text{Kendala :} \quad 0 \leq \alpha_i \leq C, \quad \sum_{i=1}^n \alpha_i y_i = 0$$

dimana α_i adalah *lagrange multiplier*, y kelas data, dan $K(x_i, x_j)$ adalah fungsi kernel sehingga diperoleh nilai *lagrange multiplier* yang optimum. Untuk melakukan prediksi kelas pada suatu data, persamaan yang digunakan adalah sebagai berikut (Fide, et al., 2021):

$$G(x) = \text{sign} \sum_{i=1}^{SVM} \alpha_i y_i K(x_i, x_j) + b \quad \dots (5)$$

dimana

$$G(x) = +1, \quad \text{jika} \quad \sum_{i=1}^{nSVM} \alpha_i y_i K(x_i, x_j) + b \geq 0,$$

$$G(x) = -1, \quad \text{jika} \quad \sum_{i=1}^{n_{SVM}} \alpha_i y_i K(x_i, x_j) + b \leq 0.$$

Dalam kasus *non-separable*, 2 kelas data tidak dapat terpisah secara sempurna. Oleh karena itu data perlu dipetakan ke suatu ruang dengan dimensi yang lebih tinggi. Fungsi kernel dapat memetakan data dari *input space* (ruang input) ke *feature space* (ruang fitur) yang berdimensi lebih tinggi agar data dapat dengan mudah dipisahkan menggunakan persamaan *hyperplane* (Cristianini & Taylor, 2014). Beberapa fungsi kernel yang biasa digunakan dalam SVM adalah sebagai berikut:

1. Kernel *Linear*

$$K(x_i, x_j) = (x_i \cdot x_j)$$

2. Kernel *Polynomial*

$$K(x_i, x_j) = ((x_i \cdot x_j) + 1)^d$$

3. Kernel *Radial Basis Function* (RBF)

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$$

4. Kernel *Sigmoid*

$$K(x_i, x_j) = \tanh(\gamma x_i \cdot x_j + r)$$

Kernel *radial basis function* (RBF) mempunyai dua parameter yaitu *cost* (C) dan γ yang harus diestimasi nilainya sehingga diperoleh parameter terbaik yang dapat secara akurat menyelesaikan kasus klasifikasi. Nilai dari parameter C dan γ tersebut dapat diestimasi menggunakan metode *cross validation*. *Cross validation* bekerja dengan memperkirakan dengan baik kesalahan prediksi atau *error rate*. Direkomendasikan untuk menggunakan metode “*grid search*” pada parameter C dan γ dalam validasi silang. Pada *grid search* berbagai pasang nilai C dan γ dicoba dan diambil pasangan nilai yang memiliki akurasi validasi silang terbaik.

Metode *one-against-one* (satu lawan satu) adalah metode yang bekerja dengan membentuk sebanyak $\frac{m(m-1)}{2}$ buah klasifikasi biner dalam penyelesaiannya dimana m adalah banyak kelas data. Sehingga, untuk 3 kelas klasifikasi, digunakan 3 buah SVM biner dalam penyelesaiannya dimana persamaan SVM-nya adalah sebagai berikut (Santosa, 2015):

Tabel 1. Ilustrasi Penggunaan Metode *One Against One*

$y_j = 1$	$y_j = -1$	Hipotesis
Kelas 1	Kelas 2	$f^{12}(x) = (w^{12})x + b^{12}$
Kelas 1	Kelas 3	$f^{13}(x) = (w^{13})x + b^{13}$
Kelas 2	Kelas 3	$f^{23}(x) = (w^{23})x + b^{23}$

Akurasi dalam klasifikasi adalah jumlah dari hasil klasifikasi yang diklasifikasikan benar sesuai dengan data aktualnya (Patro & Patra, 2014). Nilai APER (*Apparent Error Rate*) adalah persentase dari banyaknya kesalahan klasifikasi. Sehingga semakin kecil nilai APER maka ketepatan klasifikasi semakin besar (Johnson & Wichern, 2007). Dalam penelitian ini akan digunakan 3 kelas data sehingga cara menghitung tingkat akurasi dan nilai APER dapat dilakukan dengan langkah berikut,

Tabel 2. *Confusion Matrix*

Kelas	Hasil prediksi klasifikasi			Total
	1	2	3	
1	n_{11}	n_{12}	n_{13}	$n_{11} + n_{12} + n_{13}$
2	n_{21}	n_{22}	n_{23}	$n_{21} + n_{22} + n_{23}$
3	n_{31}	n_{32}	n_{33}	$n_{31} + n_{32} + n_{33}$
Total	$n_{11} + n_{21} + n_{31}$	$n_{12} + n_{22} + n_{32}$	$n_{13} + n_{23} + n_{33}$	$n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} + n_{31} + n_{32} + n_{33}$

Berdasarkan Tabel 2, Akurasi dan nilai APER dapat dihitung menggunakan persamaan berikut

$$Akurasi = \left(\frac{n_{11} + n_{22} + n_{33}}{n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} + n_{31} + n_{32} + n_{33}} \right) \times 100\%$$

$$\text{nilai APER} = \left(\frac{n_{12} + n_{13} + n_{21} + n_{23} + n_{31} + n_{32}}{n_{11} + n_{12} + n_{13} + n_{21} + n_{22} + n_{23} + n_{31} + n_{32} + n_{33}} \right) \times 100\%$$

Confusion matrix pada tabel 2 juga dapat digunakan untuk mengevaluasi hasil klasifikasi berdasarkan nilai akurasi, presisi, *recall*, dan nilai *F1-score* dengan persamaan berikut:

$$\begin{aligned} \text{Akurasi} &= \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \\ \text{Presisi} &= \frac{TP}{TP + FP} \times 100\% \\ \text{Recall} &= \frac{TP}{TP + FN} \times 100\% \\ \text{F1 - Score} &= \frac{2 \times (\text{presisi} \times \text{recall})}{\text{presisi} + \text{recall}} \times 100\% \end{aligned}$$

Akurasi merupakan penilaian terhadap tingkat keakuratan model klasifikasi. Presisi adalah rasio seberapa banyak hasil data yang relevan dengan total jumlah data yang ditemukan dari model klasifikasi. *Recall* merupakan rasio seberapa banyak data yang ditemukan kembali oleh model klasifikasi dengan total jumlah data yang relevan. (Ginantra, et al., 2022). *F1-score* dapat diartikan sebagai bobot rata-rata antara presisi dan *recall*, dimana *F1-score* mencapai nilai terbaik saat bernilai 1 dan terburuk bernilai 0 (Singgalen, 2023).

Merlinda Wibowo dan Rafian Rahmadhani telah membuktikan bahwa metode *decision tree* dan *SVM* adalah dua metode dengan tingkat akurasi tertinggi dibanding metode-metode klasifikasi lainnya (Wibowo & Ramadhani, 2021). Hiya Nalatissifa, dkk. (2020) membandingkan algoritma klasifikasi *naive bayes*, *SVM*, dan *random forest* pada kasus prediksi ketidakhadiran di tempat kerja dan menyimpulkan ketiga metode sudah dapat memprediksi ketidakhadiran dengan akurasi, presisi dan *recall* diatas 96% dan metode *SVM* dan *random forest* mencapai tingkat akurasi lebih dari 99% (Nalatissifa, et al., 2020). Bayu Sugara dan Agus Subekti menggunakan *SVM* untuk mendeteksi gangguan autisme sejak dini dengan 67 data set yang menghasilkan tingkat akurasi tertinggi sebesar 85% (Sugara & Subekti, 2019). Ana Mariyam Puspitasari, dkk. (2018) menggunakan *SVM* untuk mengklasifikasi penyakit gigi dan mulut pada manusia menggunakan 4 fungsi kernel dan diperoleh akurasi tertinggi sebesar 93.328% dengan menggunakan kernel *radial basis function* (Puspitasari, et al., 2018). Puspita Anna Octaviani, dkk. (2014) menggunakan metode *SVM multiclass* dengan membandingkan 2 fungsi kernel yaitu kernel RBF dan kernel *polynomial* dan diperoleh tingkat akurasi tertinggi menggunakan kernel RBF dengan akurasi sebesar 100% pada data *training* dan 93.902% pada data *testing* (Octaviani, et al., 2014). Beberapa penelitian sebelumnya menunjukkan bahwa *SVM* dengan kernel RBF memiliki akurasi yang tinggi dalam proses klasifikasi. Sehingga dalam penelitian ini memiliki tujuan membuat model klasifikasi tingkat keparahan kecelakaan yang terjadi di kota Padang pada bulan Januari – Maret 2022 menggunakan *SVM* dengan Kernel RBF.

2. METODE PENELITIAN

Dalam penelitian ini data yang digunakan adalah data sekunder yang diperoleh dari Kantor Satlantas Kota Padang, Sumatera Barat tentang kecelakaan lalu lintas yang dialami oleh pengendara sepeda motor yang terjadi pada bulan Januari - Maret tahun 2022. Variabel terikat (y) dari penelitian ini adalah kondisi pengendara sepeda motor setelah mengalami kecelakaan yang dinyatakan sebagai: Luka Ringan (LR), luka berat (LB), dan Meninggal Dunia (MD). Variabel bebas dari penelitian ini adalah data dari korban kecelakaan yang mempengaruhi kondisi pengendara sepeda motor setelah mengalami kecelakaan yang terdiri dari 6 jenis variabel yaitu jenis kelamin (X_1), usia (X_2), pihak lain yang ikut terlibat dalam kecelakaan (X_3), waktu terjadi kecelakaan (X_4), kendaraan yang digunakan pihak lain yang ikut terlibat dalam kecelakaan (X_5), dan tipe kecelakaan (X_6).

Tahap-tahap yang digunakan dalam studi kasus hingga memperoleh model terbaik ini adalah sebagai berikut:

1. Melakukan seleksi terhadap data kecelakaan yang telah diperoleh dari Satlantas Kota Padang yang sesuai dengan kriteria yang dibutuhkan peneliti.
2. Menginput data kecelakaan yang telah dipilih ke dalam *software Microsoft Excel* untuk selanjutnya diolah menggunakan bantuan *software R*.
3. Membagi data menjadi data *training* dan data *testing* dengan proporsi 75:25.

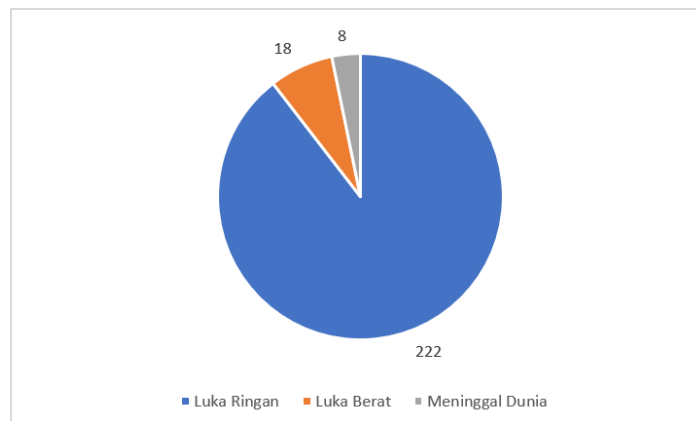
4. Menduga parameter fungsi kernel *radial basis function* (γ) dan parameter C menggunakan *K-fold cross validation* dengan metode *grid search* lakukan juga pendugaan parameter C yang dibutuhkan dalam kernel *linear*.
5. Melakukan klasifikasi SVM menggunakan kernel RBF dan *linear*.
6. Membandingkan hasil klasifikasi dengan kernel RBF dan *linear* berdasarkan Akurasi dan nilai APER pada data *training* dan *testing*.
7. menyusun model klasifikasi SVM *multiclass* berdasarkan metode *one-against-one*.

3. HASIL DAN PEMBAHASAN

Bagian ini akan dibahas tentang model klasifikasi tingkat keparahan kecelakaan sepeda motor di Kota Padang pada bulan Januari - Maret tahun 2022 menggunakan metode *support vector machine multiclass*.

3.1 Analisis Deskripsi Data

Pada gambar di bawah ini akan disajikan diagram jumlah masing-masing data berdasarkan kategorinya.



Gambar 1. Diagram Jumlah Data Berdasarkan Kategorinya

Berikut adalah analisis deskripsi data dari data training yang digunakan pada penelitian.

Tabel 3. Analisis Deskripsi Pada Data Training

Variabel	Kondisi Pengendara	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Jenis Kelamin	LB	14	1.571	0.137	0.514	1.000	1.000	2.000	2.000	2.000
	LR	166	1.729	0.035	0.446	1.000	1.000	2.000	2.000	2.000
	MD	6	1.667	0.211	0.516	1.000	1.000	2.000	2.000	2.000
Usia	LB	14	32.430	3.270	12.240	20.000	22.000	27.000	43.250	55.000
	LR	166	34.340	1.130	14.550	20.000	24.000	27.000	43.250	79.000
	MD	6	47.000	7.060	17.290	29.000	35.000	40.000	65.000	74.000
Pihak ke-2	LB	14	2.286	0.194	0.726	1.000	2.000	2.000	3.000	3.000
	LR	166	1.886	0.054	0.699	1.000	1.000	2.000	2.000	3.000
	MD	6	2.500	0.224	0.548	2.000	2.000	2.500	3.000	3.000
waktu	LB	14	1.714	0.125	0.469	1.000	1.000	2.000	2.000	2.000
	LR	166	1.765	0.033	0.425	1.000	2.000	2.000	2.000	2.000
	MD	6	1.833	0.167	0.408	1.000	1.750	2.000	2.000	2.000
kendaraan pihak ke-2	LB	14	1.429	0.137	0.514	1.000	1.000	1.000	2.000	2.000
	LR	166	1.211	0.032	0.409	1.000	1.000	1.000	1.000	2.000
	MD	6	1.500	0.224	0.548	1.000	1.000	1.500	2.000	2.000
Tipe kecelakaan	LB	14	2.000	0.296	1.109	1.000	1.000	2.000	3.000	4.000
	LR	166	2.205	0.095	1.219	1.000	1.000	2.000	3.000	4.000
	MD	6	2.833	0.543	1.329	1.000	1.750	3.000	4.000	4.000

Data *training* yang digunakan berjumlah 186 buah data, enam variabel penelitian, dan tiga kelas data. Selanjutnya akan disajikan analisis deskripsi dari data *testing*.

Tabel 4. Analisis Deskripsi Pada Data *Testing*

Variabel	Kondisi Pengendara	N	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Jenis Kelamin	LB	4	1.750	0.250	0.500	1.000	1.250	2.000	2.000	2.000
	LR	5	1.839	0.049	0.371	1.000	2.000	2.000	2.000	2.000
	MD	2	2.000	0.000	0.000	2.000	*	2.000	*	2.000
Usia	LB	4	24.250	1.310	2.630	22.000	22.250	23.500	27.000	28.000
	LR	5	32.860	1.640	12.280	20.000	22.000	28.500	42.000	63.000
	MD	2	40.000	8.000	11.310	32.000	*	40.000	*	48.000
Pihak ke-2	LB	4	2.000	0.000	0.000	2.000	2.000	2.000	2.000	2.000
	LR	5	1.875	0.092	0.689	1.000	1.000	2.000	2.000	3.000
	MD	2	2.000	1.000	1.410	1.000	*	2.000	*	3.000
waktu	LB	4	1.750	0.250	0.500	1.000	1.250	2.000	2.000	2.000
	LR	5	1.750	0.058	0.437	1.000	1.250	2.000	2.000	2.000
	MD	2	2.000	0.000	0.000	2.000	*	2.000	*	2.000
kendaraan pihak ke-2	LB	4	1.000	0.000	0.000	1.000	1.000	1.000	1.000	1.000
	LR	5	1.179	0.052	0.387	1.000	1.000	1.000	1.000	2.000
	MD	2	1.500	0.500	0.707	1.000	*	1.500	*	2.000
Tipe Kecelakaan	LB	4	2.500	0.500	1.000	1.000	1.500	3.000	3.000	3.000
	LR	5	2.018	0.145	1.087	1.000	1.000	2.000	3.000	4.000
	MD	2	1.000	0.000	0.000	1.000	*	1.000	*	1.000

Penelitian ini menggunakan total 248 buah data dimana 25% diantaranya dijadikan data *testing* atau sebanyak 62 buah data, dengan statistik deskripsi data dapat dilihat pada Tabel 4.

3.2 Nilai Parameter Terbaik Fungsi Kernel

Dalam klasifikasi menggunakan menggunakan *SVM*, terdapat beberapa parameter yang harus ditentukan. Diantaranya parameter *cost* (C) dan parameter *gamma* (γ) pada fungsi kernel *radial basis function* (RBF). Metode *K-fold cross validation* (validasi silang) dapat digunakan untuk menentukan variabel nilai dari parameter C dan γ dengan bantuan metode *grid search*. Metode ini bekerja dengan mencoba beberapa nilai parameter yang selanjutnya dengan bantuan *software R studio* akan dipilih satu nilai parameter terbaik untuk digunakan dalam klasifikasi menggunakan *SVM*. Dalam penelitian ini, untuk parameter C akan digunakan parameter 2^{-2} sampai 2^{12} dan untuk parameter γ digunakan parameter 2^{-10} sampai 2^4 . Dengan menggunakan *software R studio*, diperoleh parameter terbaik untuk parameter C adalah $2^0 = 1$ dan parameter γ adalah $2^0 = 1$. Sebagai pembandingan hasil klasifikasi menggunakan kernel RBF, akan dilakukan klasifikasi lain menggunakan kernel *linear*, dipilihnya kernel *linear* sebagai pembandingan karena kernel RBF bekerja dengan mengubah dimensi data dari *input space* ke *feature space* secara non linier sedangkan kernel *linear* mengubah dimensi data secara linier. Pada kernel *linear* terdapat satu parameter yang perlu diestimasi yaitu parameter *cost* (C), dengan menggunakan rentang yang sama dengan yang digunakan pada kernel RBF yaitu 2^{-2} sampai 2^{12} diperoleh estimasi nilai parameter *cost* sebesar $2^{-2} = 0.25$.

3.3 Ketepatan Klasifikasi Fungsi Kernel

Menggunakan parameter yang sudah diperoleh dapat dilakukan klasifikasi dengan ketepatan klasifikasi berdasarkan akurasi dan APER menggunakan fungsi kernel RBF dan *linear* pada data *training* dan *testing* dapat dilihat pada tabel berikut.

Tabel 5. Hasil Klasifikasi Data *Training* Menggunakan Kernel RBF

Data aktual	Hasil prediksi klasifikasi			Total	Akurasi	APER
	LR	LB	MD			
LR	166	0	0	166	94.62%	5.38%
LB	8	6	0	14		
MD	2	0	4	6		
Total	166	6	4	186		

$$Akurasi = \frac{166 + 6 + 4}{166 + 0 + 0 + 8 + 6 + 0 + 2 + 0 + 4} \cdot 100\% = 94.62\%$$

$$Nilai APER = \frac{0 + 0 + 8 + 0 + 2 + 0}{166 + 0 + 0 + 8 + 6 + 0 + 2 + 0 + 4} \cdot 100\% = 5.38\%$$

Tabel 6. Hasil Klasifikasi Data *Training* Menggunakan Kernel *Linear*

Data aktual	Hasil prediksi klasifikasi			Total	Akurasi	APER
	LR	LB	MD			
LR	166	0	0	166	89.25%	10.75%
LB	14	0	0	0		
MD	6	0	0	0		
Total	186	0	0	186		

$$Akurasi = \frac{166 + 0 + 0}{166 + 0 + 0 + 14 + 0 + 0 + 6 + 0 + 0} \cdot 100\% = 89.25\%$$

$$nilai APER = \frac{0 + 0 + 14 + 0 + 6 + 0}{166 + 0 + 0 + 14 + 0 + 0 + 6 + 0 + 0} \cdot 100\% = 10.75\%$$

Berdasarkan hasil klasifikasi yang dilakukan pada data *training* menggunakan kernel RBF dan kernel *linear*. Diperoleh hasil klasifikasi menggunakan kernel RBF dengan tingkat akurasi sebesar 94.62% dan nilai APER sebesar 5.28%. sedangkan pada kernel *linear* diperoleh akurasi sebesar 89.25% dan nilai APER sebesar 10.75% dimana tingkat akurasi menggunakan kernel RBF lebih besar serta mempunyai nilai APER yang lebih kecil dibandingkan menggunakan kernel *linear*. Hal ini menunjukkan bahwa, untuk data kecelakaan yang digunakan pada penelitian ini, diperoleh bahwa klasifikasi menggunakan kernel RBF lebih baik daripada menggunakan kernel *linear*. Berikut ini adalah hasil dari perhitungan akurasi, presisi, *recall*, dan *F1-score* pada masing-masing kernel untuk data *training*.

Tabel 7. Akurasi, Presisi, *Recall* dan *F1-Score* dari Masing-Masing Kernel pada Data *Training*.

Kategori	Kernel RBF				Kernel Linear			
	Akurasi	Presisi	<i>Recall</i>	<i>F1-score</i>	Akurasi	Presisi	<i>Recall</i>	<i>F1-score</i>
LR	94,62%	94,32%	100%	97,07%	89,25%	89,25%	100%	94,17%
LB	90,30%	42,80%	100%	59,90%	94,62%	0%	0%	0%
MD	98,8%	66%	100%	80%	96,77%	0%	0%	0%

Berdasarkan hasil perhitungan yang dapat dilihat pada tabel 7. *F1-score* untuk kernel RBF pada kategori luka ringan sebesar 97,07%, luka berat 59,90% dan meninggal dunia sebesar 80%. Sedangkan pada penggunaan kernel *linear* diperoleh *F1-score* dari kategori luka ringan sebesar 94,17%, luka berat dan meninggal dunia masing-masing 0% karna pada *confusion matrix* tidak

terdapat data yang diprediksi masuk kedalam dua kategori tersebut. Berdasarkan kriteria dari *F1-score* dapat kita simpulkan bahwa penggunaan kernel RBF jauh lebih baik dibandingkan penggunaan kernel *linear*.

Tabel 8. Ketepatan Klasifikasi SVM pada Data *Testing* dengan Kernel RBF

Data Aktual	Hasil prediksi klasifikasi			Total	Akurasi	APER
	LR	LB	MD			
LR	56	0	0	56	90,32%	9,68%
LB	4	0	0	4		
MD	2	0	0	2		
Total	62	0	0	62		

$$\text{Akurasi} = \frac{56 + 0 + 0}{56 + 0 + 0 + 4 + 0 + 0 + 2 + 0 + 0} \cdot 100\% = 90.32\%$$

$$\text{Nilai APER} = \frac{0 + 0 + 4 + 0 + 2 + 0}{56 + 4 + 2 + 0 + 0 + 0 + 0 + 0 + 0} \cdot 100\% = 9.68\%$$

Tabel 9. Ketepatan Klasifikasi SVM pada Data *Testing* dengan Kernel *Linear*

Data Aktual	Hasil prediksi klasifikasi			Total	Akurasi	APER
	LR	LB	MD			
LR	56	0	0	56	90,32%	9,68%
LB	4	0	0	4		
MD	2	0	0	2		
Total	62	0	0	62		

$$\text{Akurasi} = \frac{56 + 0 + 0}{56 + 0 + 0 + 4 + 0 + 0 + 2 + 0 + 0} \cdot 100\% = 90.32\%$$

$$\text{Nilai APER} = \frac{0 + 0 + 4 + 0 + 2 + 0}{56 + 4 + 2 + 0 + 0 + 0 + 0 + 0 + 0} \cdot 100\% = 9.68\%$$

Tabel 8 dan Tabel 9 menunjukkan bahwa hasil perbandingan hasil klasifikasi menggunakan model SVM dengan data aktualnya menggunakan kernel RBF dan *linear*. Pada tabel terlihat bahwa ketepatan klasifikasi berdasarkan Akurasi dan nilai APER dari kedua kernel yang dibandingkan sama yaitu dengan akurasi sebesar 90.32% dan APER sebesar 9.68% dalam mengklasifikasi data *testing*.

Tabel 10. Akurasi, presisi, *recall* dan *F1-score* dari masing-masing kernel pada data *testing*.

Kategori	Kernel <i>linear</i>				Kernel RBF			
	Akurasi	Presisi	<i>Recall</i>	<i>F1-score</i>	Akurasi	Presisi	<i>Recall</i>	<i>F1-score</i>
LR	90,32%	90,32%	100%	94,91%	90,32%	90,32%	100%	94,91%
LB	96,77%	0%	0%	0%	96,77%	0%	0%	0%
MD	93,54%	0%	0%	0%	93,54%	0%	0%	0%

Pada Tabel 10, dapat dilihat bahwa pada data *testing* terdapat nilai *F1-score* di 2 kategori data yang bernilai 0% untuk kedua fungsi kernel yang digunakan. Hal ini menandakan buruknya nilai presisi dan *recall* yang dimiliki kategori luka berat dan meninggal dunia. hal ini diakibatkan

tidak ada data yang terklasifikasi kedalam kategori luka berat dan meninggal dunia sedangkan data aktual mencatat adanya kasus kecelakaan yang mengalami dua kategori tersebut. hal ini disebabkan oleh jumlah data pada setiap kategori yang tidak seimbang (*imbalanced class data*) sehingga model condong untuk memprediksi pada kelas data yang jumlahnya lebih banyak dibandingkan kelas data yang jumlah sedikit. Sehingga hasil *F1-score* akan bagus pada data mayoritas dan buruk pada data minoritas (Chawla, 2003).

3.4 Model Klasifikasi SVM

SVM dapat diselesaikan dengan bantuan *lagrange multiplier* dimana dengan mensubstitusi data pada persamaan *lagrange multiplier* menggunakan fungsi kernel RBF dengan nilai parameter γ yang telah diperoleh sebelumnya dapat ditentukan nilai masing-masing α dengan syarat $0 \leq \alpha_i \leq C$. Pada kasus kecelakaan yang diteliti, diperoleh sebanyak 158 *number of support vector* dimana *number of support vector* adalah entri data yang menjadi titik *support vector* dan memiliki nilai α yang memenuhi untuk disubstitusi ke dalam persamaan klasifikasi. Dengan nilai α yang telah didapat, dapat ditentukan nilai parameter b yaitu 0.9228773 untuk kasus 1, 0.9526409 untuk kasus 2 dan 0.52104171 untuk kasus 3. Menggunakan kernel RBF dengan metode SVM *multiclass one-against-one* dapat dibentuk persamaan klasifikasi SVM *multiclass* dengan 3 kelas data sebagai berikut:

untuk kasus 1, dimana LR (+1) dan LB (-1):

$$G^1(x) = \sum_{i=1}^{139} \alpha_i y_i \exp \exp (-1 \|x_i - x_j\|^2) + 0.9228773,$$

untuk kasus 2, dimana LR (+1) dan MD (-1):

$$G^2(x) = \sum_{i=1}^{142} \alpha_i y_i \exp \exp (-1 \|x_i - x_j\|^2) + 0.9526409,$$

untuk kasus 3, dimana LB (+1) dan MD (-1):

$$G^3(x) = \sum_{i=1}^{19} \alpha_i y_i \exp \exp (-1 \|x_i - x_j\|^2) + 0.5204171.$$

4. SIMPULAN DAN SARAN

Dengan membandingkan penggunaan dua fungsi kernel yaitu kernel *radial basis function* (RBF) dan kernel *linear* pada SVM *multiclass* dalam klasifikasi tingkat keparahan kecelakaan lalu lintas kota Padang bulan Januari - Maret tahun 2022, diperoleh bahwa kernel RBF dapat mengklasifikasi lebih baik pada data *training* dengan tingkat akurasi sebesar 94.62% dan nilai APER sebesar 5.38%. hasil klasifikasi data *training* dengan kernel RBF juga menghasilkan *F1-score* sebesar 97,07% pada kategori luka ringan, 59,90% pada kategori luka berat dan 80% pada kat. Dengan mengestimasi nilai parameter yang diperlukan dalam melakukan klasifikasi menggunakan kernel RBF digunakan metode *K-fold cross validation*, diperoleh nilai parameter $C = 1$ dan parameter adalah $\gamma = 1$. Sehingga dapat dibentuk model klasifikasi SVM *multiclass* sebagai berikut:

untuk kasus 1, dimana LR (+1) dan LB (-1):

$$G^1(x) = \sum_{i=1}^{139} \alpha_i y_i \exp \exp (-1 \|x_i - x_j\|^2) + 0.9228773,$$

untuk kasus 2, dimana LR (+1) dan MD (-1):

$$G^2(x) = \sum_{i=1}^{142} \alpha_i y_i \exp \exp (-1 \|x_i - x_j\|^2) + 0.9526409,$$

untuk kasus 3, dimana LB (+1) dan MD (-1):

$$G^3(x) = \sum_{i=1}^{19} \alpha_i y_i \exp \exp (-1 \|x_i - x_j\|^2) + 0.5204171,$$

dengan *number of support vector* sebanyak 158 dari 186 data *training* yang digunakan. Untuk penelitian selanjutnya disarankan menggunakan *SVM multiclass* yang lebih dari 3 kelas data atau membandingkan *SVM* dengan metode klasifikasi lain seperti *decision tree*, regresi logistik, *naïve bayes*, jaringan syaraf tiruan dan lain sebagainya.

DAFTAR PUSTAKA

- Chawla, N. V., 2003. C4.5 and Imbalanced Data Sets: Investigating the effect of sampling method, probabilistic estimate, and decision tree structure. *ICML Workshop Learning from Imbalanced Data Sets II. Washington D.C.*
- Cristianini, N. & Taylor, J. S., 2014. *An Introduction of Support Vector Machine and Other Kernel-Based Learning Method*. United Kingdom: Cambridge University Press.
- Enggarsari, U. & Sa'diyah, N. K., 2017. Kajian Terhadap Faktor-Faktor Penyebab Kecelakaan Lalu Lintas dalam Upaya Perbaikan Pencegahan Kecelakaan Lalu Lintas. *Perspektif*, **22**(3), pp. 238-247.
- Fide, S., Suparti & Sudarno, 2021. Analisis Sentimen Ulasan Aplikasi Tiktok di Google Play Menggunakan Support Vector Machine (SVM) dan Asosiasi. *Jurnal Gaussian*, **10**(3), pp. 346-358.
- Ginantra, N. L. W. S. R. et al., 2022. Analisis Setimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-nn. *Janapati*, **11**(3), pp. 206-216.
- Hastie, T., Tibshirani, R. & Friedman, J., 2008. *The Element of Statistical Learning*. 2nd ed. Stanford: Spinger Science Business Media.
- Johnson, R. A. & Wichern, D. W., 2007. *Applied Multivariate Statistical Analysis*. 6nd ed. United States of America: Pearson Education.
- Kusumahadi, S. H., Junaedi, H. & Santoso, J., 2019. Klasifikasi Helpdesk Menggunakan Metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT (JPIT)*, **4**(1), pp. 55-60.
- Nalatissifa, H., Gata, W., Diantika, S. & Nisa, K., 2020. Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja. *Jurnal Informatika Universitas Pamulang*, **5**(4), pp. 578-584.
- Octaviani, P. A., Wilandari, Y. & Ispriyanti, D., 2014. Penerapan Metode Klasifikasi Support Vector Machine (SVM) pada Data Akreditasi Sekolah Dasar (SD) di Kabupaten Magelang. *Jurnal Gaussian*, **3**(4), pp. 811-820.
- Patro, V. M. & Patra, M. R., 2014. Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy. *Society for Science and Education*, **2**(4), pp. 77-91.
- Puspitasari, A. M., Ratnawati, D. E. & Widodo, A. W., 2018. Klasifikasi Penyakit Gigi Dan Mulut Menggunakan Metode Support Vector Machine. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, **2**(2), pp. 802-810.
- Rachman, F. & Purnami, S. W., 2012. Perbandingan Klasifikasi Tingkat Keganasan Breast Cancer dengan Menggunakan Regresi Logistik Ordinal dan Support Vector Machine (SVM). *Jurnal Sains dan Seni ITS*, **1**(1), pp. 130-135.
- Santosa, B., 2015. Multiclass Classification with Cross Entropy-Support Vector. *Propedia Computer Sains*, Volume **72**, pp. 345-352.
- Scholkopf, B. & Smola, A. J., 2002. *Learning With Kernel*. England: Massachusetts Institute of Technology.
- Sugara, B. & Subekti, A., 2019. Penerapan Support Vector Machine (SVM) pada Small Dataset untuk Deteksi Dini Gangguan Autisme. *Jurnal PILAR Nusa Mandiri*, **15**(2), pp. 177-182.
- Wibawa, A. P., Purnama, M. G. A., Akbar, M. F. & Dwiyanto, F. A., 2018. Metode-Metode Klasifikasi. *Prosiding Seminar Ilmu Komputer dan Teknologi Informasi*, **3**(1), pp. 134-138.
- Wibowo, M. & Ramadhani, R., 2021. Perbandingan Metode Klasifikasi Data Mining Untuk Rekomendasi Tanaman Pangan. *Jurnal Media Informatika Budidarma*, **5**(1), pp. 913-921.