



Klasifikasi Kategori Pertandingan Atlet Silat Perisai Diri Menggunakan Metode Gaussian Naïve Bayes Berbasis WEB

Dwi Saka Pangestu^{1*}, Agung Nilogiri², Deni Arifianto³

^{1,2,3}Program Studi Teknik Informatika, Universitas Muhammadiyah Jember

Email: sakaestu@gmail.com^{1*}, agungnilogiri@unmuhjember.ac.id², deniarifianto@unmuhjember.ac.id³

ABSTRAK

Perisai Diri merupakan salah satu dari pencak silat yang ada di Indonesia dan aktif mengadakan kejuaraan internal atau kejuaraan antar unit/ranting Perisai Diri itu sendiri. Pertandingan internal Perisai Diri memiliki tiga kategori yaitu Tanding (Fight), TGR (Tunggal, Ganda, Regu)/Seni, dan Serang Hindar, dimana dari kategori pertandingan tersebut memiliki karakteristik dan kebutuhan pertandingan yang berbeda. Perisai Diri Jember masih belum memiliki sistem untuk membantu memilih atlet, selama ini pemilihan kategori untuk atlet dilakukan secara manual. Tujuan dari penelitian ini yaitu untuk mengetahui tingkat akurasi, presisi, dan recall dalam klasifikasi atlet Perisai Diri dengan menggunakan 64 data. Metode klasifikasi yang digunakan pada penelitian kali ini yaitu Gaussian Naïve Bayes dan menggunakan metode pendukung K-Fold Cross Validation. Hasil dari penelitian ini didapatkan akurasi sebesar 57,1%, dengan recall terbesar terdapat pada TGR 80% dan presisi terbesar terdapat pada Fight 60%.

Kata Kunci: Klasifikasi Atlet, Gaussian Naïve Bayes, Pencak Silat

ABSTRACT

Perisai Diri is one of the existing pencak silat in Indonesia and actively holds an internal championships or championships between units/branches of Perisai Diri itself. Perisai Diri's internal matches have three categories, Tanding (Fight), TGR (Tunggal, Ganda, Regu)/Seni, and Serang Hindar, where each of these competition categories has a different characteristics and needs for matches. Perisai Diri Jember still doesn't have a system to help select athletes, so far the selection of categories for athletes was manually. The purpose of this study was to measure the level of accuracy, precision, and recall in classification of Perisai Diri athletes using 64 data. The classification method used in this study is Gaussian Naïve Bayes by using the support method K-fold Cross Validation. The result of this study obtained accuracy 57,1%, with the largest recall in TGR 80% and the greatest precision in Fight 60%..

Keywords: Athletes Classification, Gaussian Naïve Bayes, Pencak Silat

1. PENDAHULUAN

Pencak silat adalah sebuah seni bela diri dan olahraga asli Indonesia yang perlu dikembangkan oleh seluruh masyarakat dari anak-anak hingga orang dewasa. Antusiasme masyarakat terhadap olahraga pencak silat menjadikan pencak silat sebagai salah satu cabang olahraga yang dipertandingkan di kancah regional, nasional dan internasional. Pencak silat di Indonesia memiliki beberapa cabang perguruan tinggi yang tersebar di berbagai daerah di Indonesia dengan IPSI (Ikatan Pencak Silat Indonesia) sebagai induk dari semua aliran pencak silat di Indonesia. Adapun 10 organisasi pencak silat memegang peranan penting dalam organisasi induk Pencak Silat Indonesia PD (Perisai Diri), Persaudaraan Setia Hati, Persaudaraan Setia Hati Terate, Tapak Suci, Phasadja Mataram, Persatuan Pencak Silat Indonesia, PPS Putra Betawi dan lain-lain (Pratama, Rendra & Trilaksana, 2018).

Keluarga Silat Nasional (Kelatnas) Perisai Diri Jember merupakan salah satu organisasi beladiri yang selalu ikut serta dalam mengikuti pertandingan pencak silat baik dalam kancah IPSI (pertandingan antar perguruan) maupun pertandingan antar unit/ranting Perisai Diri itu sendiri (Internal). Pertandingan internal Perisai Diri memiliki tiga (3) kategori, yaitu Tanding, Serang Hindar, dan Seni TGR (Tunggal, Ganda, Regu). Dimana dari kategori pertandingan tersebut memiliki karakteristik dan kebutuhan pertandingan yang berbeda. Sebagai contoh atlet kategori tanding dan

TGR memiliki kebutuhan kekuatan (Power) dan daya tahan (Stamina) yang berbeda. Dari hasil diskusi dengan pengurus dan pelatih, kelatnas Perisai Diri Jember masih belum memiliki sistem untuk membantu memilih atlet, selama ini pemilihan kategori untuk atlet dilakukan dengan cara manual, sehingga memakan waktu yang cukup lama. Dengan adanya permasalahan tersebut, akan sangat membantu jika menciptakan sistem pendukung keputusan pada menentukan kategori atlet silat. Sistem pendukung keputusan merupakan sistem yang dipakai untuk membantu membuat atau memilih keputusan, diantara banyak keputusan yang didapatkan menurut pemrosesan informasi yang tersedia (Hutahaean & Hasiguan, 2021). Untuk proses pembuatan sistem pendukung keputusan perlu untuk memperhatikan beberapa hal yaitu beberapa klasifikasi untuk menentukan keputusan kategori pertandingan apa yang cocok untuk atlet silat.

2. KAJIAN PUSTAKA

A. Pencak Silat

Seni bela diri asal Indonesia yang memiliki gerakan yang indah, baik dalam melangkah, menyerang, dan menghindari adalah Pencak Silat. Pencak dan Silat memiliki makna tersendiri yang mana Pencak dapat di pertandingkan dalam ajang prestasi, sedangkan Silat merupakan bela diri yang tidak di peruntukkan sebagai tontonan masyarakat, melainkan hanya sebagai bentuk pembelaan diri, karena di dalamnya terdapat serangan, hindaran dan teknik mengunci (Maryono, 1998).

B. Sistem Pendukung Keputusan

Menurut Turban, sistem pendukung keputusan (SPK) merupakan sistem berbasis model yang terdiri dari prosedur pemrosesan data dan inferensi yang membantu manajer dalam pengambilan keputusan. Untuk mencapai tujuan itu, sistem saat tersebut harus sederhana, mudah dikendalikan, dapat disesuaikan, dan lengkap. Sistem ini digunakan untuk mendukung pengambilan keputusan dalam situasi semi terstruktur dan tidak terstruktur di mana tidak ada yang tahu pasti bagaimana keputusan harus dibuat (Fauzi, 2016).

C. Naïve Bayes

Menurut Bustami (2014) Thomas Bayes merupakan ilmuwan inggris yang mengusulkan metode probabilistic dan statistik yang kemudian di kenal dengan teorema Bayes, yang mana teorema ini dapat memprediksi peluang masa depan berdasarkan pengalaman masa lalu. Teorema ini lalu digabungkan dengan Naive dan diasumsikan bahwa ada atau tidak ciri tertentu dari sebuah kelas tidak ada hubungannya dengan ciri kelas lainnya. Persamaan dari teorema bayes adalah (Bustami, 2014):

$$P(H|X) = \frac{P(X|H)}{P(X)} \cdot P(H) \quad (1)$$

Dimana:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatau class spesifik

P(H|X) = Probabilitas hipotesis H berdasar kondisi X (Posteriori probability)

P(H) = Probabilitas.hipótesis H (Prior Probability)

Jika menyangkut data kontinu maka akan menggunakan rumus Densitas gauss (Bustami, 2014):

$$P(X_i = x_i|Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}}} e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (2)$$

Dimana:

P = Peluang

X_i = Atribut ke i

x_i = Nilai atribut ke i

Y = Kelas yang dicari

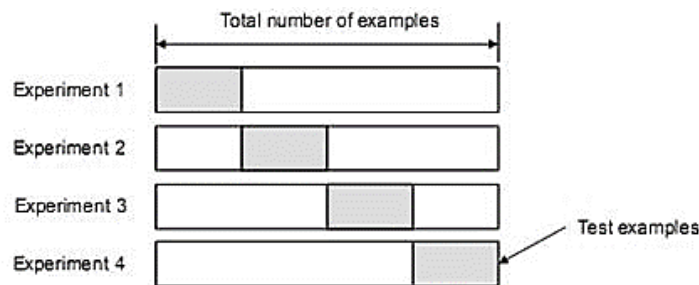
y_j = Sub kelas Y yang dicari

μ = Mean, menyatakan rata rata dari seluruh atribut

σ = Deviasi standar, menyatakan varian dari seluruh atribut

D. K-Fold Cross Validation

K-Fold Cross validation ialah satu kategori *cross validation* yang berfungsi untuk mengukur serta memperkirakan kinerja model prediktif serta analisa statistik dengan teknik membagi data sample secara acak jadi K *sub-sample* sebanyak nilai K . Setelah itu, satu *sub-sampel* dijadikan sebagai data validasi untuk pengujian model, dan *sub-sample* $K-1$ yang tersisa dipakai sebagai data pelatihan. Proses ini diulang sebanyak K kali (Rohani et al., 2018). Dimisalkan dengan $K_f = 4$, dimana data hendak dipecah sebanyak 4 himpunan. Pada iterasi pertama berisi 3 himpunan data yaitu D_2, D_3, D_4 sebagai data latih serta satu himpunan D_1 untuk data uji. Pada iterasi kedua berisi 3 himpunan data yaitu: D_1, D_3, D_4 sebagai data latih serta satu himpunan D_2 untuk data uji. Demikian seterusnya untuk iterasi ketiga dan keempat hingga setiap himpunan data pernah menjadi data uji sebanyak satu kali. Contoh penerapan *K-Fold Cross Validation* dengan nilai $K_f = 4$ yang bisa dilihat pada Gambar 1.



Gambar 1. 4 Fold Cross Validation (Sumber: Slideplayer.info)

E. Confusion Matrix

Confusion matrix ialah metode yang bisa dipakai memperkirakan kinerja prosedur klasifikasi. Hasil dari confusion matrix memiliki informasi yang bisa digunakan untuk membandingkan hasil klasifikasi yang dicoba oleh program dengan hasil klasifikasi yang sebenarnya (Rustam & Audia Ariantari, 2018).

Dalam penggunaan *Confusion Matrix* untuk mengukur kinerja metode klasifikasi terdapat 4 istilah yang mereprestasikan hasil dari klasifikasi. Berikut keempat representasi hasil proses klasifikasi yang ditunjukkan pada Tabel 1.

Tabel 1. Representasi Hasil Proses Klasifikasi *Confusion Matrix*

<i>True Positive</i> (TP)	Total data positif yang terklasifikasi dengan benar oleh sistem
<i>True Negative</i> (TN)	Total data negatif yang terklasifikasi dengan benar oleh sistem
<i>False Negative</i> (FN)	Total data positif akan tetapi terklasifikasi sebagai data negatif oleh sistem

<i>False Positive</i> (FP)	Total data negatif akan tetapi terklasifikasi sebagai data positif oleh sistem
----------------------------	--

Sumber : (Santra & Christy, 2012)

Untuk mengevaluasi suatu model klasifikasi dapat menggunakan akurasi, presisi dan *recall*. Akurasi adalah perbandingan jumlah prediksi yang benar dengan total prediksi. Presisi adalah perbandingan prediksi positif yang benar dengan total prediksi positif. Sedangkan Recall adalah persentase prediksi positif dengan data positif (Santra & Christy, 2012). Berikut merupakan rumus akurasi, presisi dan *recall*:

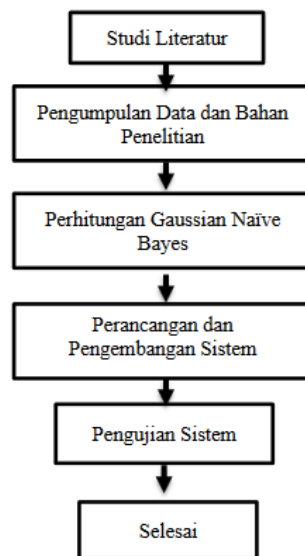
$$\text{Akurasi} = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (3)$$

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\% \quad (4)$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\% \quad (5)$$

3. METODE PENELITIAN

Dalam proses penelitian terdapat susunan alur yang terdiri dari beberapa tahap yang bertujuan untuk mempermudah proses penelitian. Berikut merupakan alur tahapan penelitian yang ditunjukkan pada Gambar 2.



Gambar 2. Tahapan penelitian

A. Studi Literatur

Studi literatur dilakukan dengan cara mencari informasi dengan membaca buku atau literatur yang berkaitan dengan penelitian ini di mana menggunakan metode yang sama yaitu *gaussian naïve bayes*.

B. Pengumpulan Data

Dalam penelitian ini pengumpulan data berupa dokumen, *file*, dan catatan – catatan diperoleh dari sekretaris Perisai Diri Jember. Dari data-data yang telah dikumpulkan maka didapatkan *dataset*. *Dataset* berupa data atlet yang telah melakukan seleksi dan mendapatkan nilai untuk ditentukan hasil kategori pertandingan silat. *Dataset* ini terdiri dari 64 data dengan 8 atribut di antaranya: nama, *speed*, *power*, *stamina*, *agility*, kedisiplinan, gerak/Teknik, dan kategori pertandingan. Dalam penelitian ini menggunakan metode *gaussian naïve bayes* untuk mengetahui hasil dari akurasi, presisi dan *recall*. Pada contoh perhitungan bab ini menggunakan 10 data latihan dan 10 data uji untuk proses perhitungan.

C. Perhitungan *Gaussian Naïve Bayes*

Untuk melakukan proses perhitungan *gaussian naïve bayes* maka inputan dari 6 nilai kriteria harus terisi di antaranya nilai *speed*, nilai *power*, nilai *stamina*, nilai *agility*, nilai kedisiplinan, dan nilai gerak/Teknik. Data latihan dapat dilihat pada Tabel 2.

Tabel 2. Data Latihan *Gaussian Naïve Bayes*

No	Nama	S	P	S	A	K	G/T	Hasil
1	Fatur	77	75	71	70	67	66	Fight
2	Budi	68	65	62	61	59	59	Fight
3	Arf	60	69	61	68	69	59	Fight
..								
10	Intan	62	58	60	57	62	62	SerHin

Berdasarkan tabel diatas kolom S,P,S,A,K,G/T adalah atribut yang mana mewakili *Speed*, *Power*, *Stamina*, *Agility*, *Kedisiplinan*, *Gerak/Teknik*. Dan pada kolom hasil terdapat nilai SerHin yang mana mewakili kategori Serang Hindar

Tabel 3. Data Uji *Gaussian Naïve Bayes*

No	Nama	S	P	S	A	K	G/T	Hasil
1	Kia	73	70	67	65	63	62	Fight
2	Abdur	72	69	66	65	63	62	Fight
3	Robi	65	68	62	69	58	60	Fight
..								
10	Eka	72	69	66	70	69	68	SerHin

Berdasarkan Tabel 3 di atas terdapat 10 data yang mana 4 data *fight* 4 data, serang hindar 2 data, dan TGR 4 data. Langkah berikutnya setelah data latihan dan data uji telah di ketahui maka menentukan probabilitas kelas

$$p(\text{Fight}) = \frac{4}{10} = 0,4$$

Berikut merupakan hasil probabilitas dari setiap kategori, pada Tabel 4 di bawah ini:

Tabel 4. Hasil Probabilitas Kelas

Kategori Pertandingan	Nilai Probabilitas
Fight	0,40
TGR	0,40
Serang Hindar	0,20

Setelah mendapatkan hasil probabilitas setiap kelas langkah berikutnya yaitu mencari nilai *mean* / rata – rata dari setiap atribut untuk masing – masing kelas

$$\mu = \frac{75 + 65 + 69 + 67}{4} = \frac{276}{4} = 69$$

Contoh di atas merupakan perhitungan dari kelas *Fight* atribut *Power*, lakukan perhitungan yang sama dengan menggunakan kelas dan atribut yang lainnya.

Tabel 5. Hasil Mean / Rata - Rata

	Speed		Power		Stamina	
	Fight	68,75	Fight	69	Fight	64,5
	TGR	64,75	TGR	71,75	TGR	70,75
	SerHin	63,5	SerHin	59,75	SerHin	63,5
	Agility		Kedisiplinan		Gerak/Teknik	
	Fight	65,5	Fight	64	Fight	61
	TGR	61,75	TGR	61,25	TGR	68,25
	SerHin	64,5	SerHin	65,75	SerHin	64,75

Tabel di atas merupakan hasil dari keseluruhan *mean* / rata – rata atribut setiap kelas. Setelah diketahui *mean* dari atribut setiap kelas maka selanjutnya melakukan pencarian nilai standard deviasi menggunakan data uji dan hasil dari *mean*. Contoh perhitungan menggunakan atribut *Speed* pada kelas *Fight*.

$$\sigma = \sqrt{\frac{(77 - 68,75)^2 + (68 - 68,75)^2 + (60 - 68,75)^2 + (70 - 68,75)^2}{4 - 1}} = \sqrt{\frac{146,74}{3}} = \sqrt{48,91} = 6,994045086$$

Pada tabel di bawah ini merupakan standard deviasi dari semua kelas pada setiap atribut.

Tabel 6. Hasil Standard Deviasi

	Speed		Power		Stamina	
	Fight	6,99	Fight	4,32	Fight	4,50
	TGR	14,8	TGR	3,20	TGR	0,95
	SerHin	2,12	SerHin	2,82	SerHin	4,94
	Agility		Kedisiplinan		Gerak/Teknik	
	Fight	4,20	Fight	4,76	Fight	3,36
	TGR	14,56	TGR	7,22	TGR	5,79
	SerHin	10,6	SerHin	5,65	SerHin	4,24

Langkah berikutnya yaitu menggunakan data uji untuk melakukan perhitungan menggunakan rumus gaussian naive bayes. Contoh menggunakan data uji ke -1 pada atribut *speed* kategori *fight*.

$$p(x) = \frac{1}{\sqrt{2} \times 3,14 \times 6,994} e^{-\frac{(73-68,75)^2}{2(6,994^2)}} = 0,1255$$

Keseluruhan data uji di hitung menggunakan rumus di atas kemudian mengalikan nilai setiap atribut dan nilai probabilitas:

$$0.1255 \times 0.1869 \times 0.1611 \times 0.1932 \times 0.1789 \times 0.2080 \times 0,40 = 1,087425E - 05$$

Pada contoh di atas merupakan nilai klasifikasi untuk kelas *Fight*, yang di dapatkan dari mengalikan hasil probabilistik dengan nilai probabilitas setiap kelas. Di bawah ini merupakan hasil klasifikasi dengan memilih nilai terbesar dari setiap kategori.

Tabel 7. Hasil Klasifikasi

Nama	Hasil	Prediksi	Real	Kriteria
Kia	1,08E-05	1	1	True
Rosyid	1,33E-05	1	1	True
Robi	3,72E-06	1	1	True
Ifan	2,81E-07	1	1	True
Meta	5,74E-08	3	2	False
Yus	2,12E-07	2	2	True
Dian	4,88E-06	1	2	False

Sofi	6,23E-08	2	2	True
Ali	5,05E-06	1	3	False
Fajar	5,35E-07	1	3	False

Berdasarkan hasil pada Tabel 7 di atas, untuk mendapatkan nilai akurasi, recall, dan presisi menggunakan *Confusion Matrix*. Yang mana recall dan presisi akan dihitung per kategori, sedangkan untuk akurasi akan dihitung secara keseluruhan. Dari tabel 4 di atas terdapat 4 *False* sehingga dapat dihitung akurasi sebagai berikut:

$$Akurasi = \frac{6}{10} \times 100\% = 60\%$$

Sedangkan untuk recall dan presisi akan menggunakan *Confusion Matrix* berikut ini:

Tabel 8. Hasil Confusion Matrix Fight

No	Prediksi	Real	Kriteria
1	1	1	TP
2	1	1	TP
3	1	1	TP
4	1	1	TP
5	3	2	TN
6	2	2	TN
7	1	2	FN
8	2	2	TN
9	1	3	FN
10	1	3	FN

Berdasarkan Tabel 8 di atas presisi dan recall dari kelas *Fight* adalah 57,14% untuk recall dan 100% untuk presisi.

Tabel 9. Hasil Confusion Matrix Fight

No	Prediksi	Real	Kriteria
1	1	1	TN
2	1	1	TN
3	1	1	TN
4	1	1	TN
5	3	2	FP
6	2	2	TP
7	1	2	FP
8	2	2	TP
9	1	3	TN
10	1	3	TN

Berdasarkan Tabel 9 di atas presisi dan recall dari kelas *TGR* adalah 100% untuk *recall* dan 75% untuk presisi.

Tabel 10. Hasil Confusion Matrix Fight

No	Prediksi	Real	Kriteria
1	1	1	TN
2	1	1	TN
3	1	1	TN
4	1	1	TN
5	3	2	FN
6	2	2	TN
7	1	2	TN
8	2	2	TN
9	1	3	FP
10	1	3	FP

Berdasarkan Tabel 10 di atas presisi dan recall dari kelas TGR adalah 0% untuk recall dan 0% untuk presisi.

4. HASIL DAN PEMBAHASAN

A. Pembagian Data

Dari 64 data atlet akan dibagi menjadi 50 data dan 14 *unseen* data test. Yang mana 50 data tersebut akan diuji menggunakan *K-Fold Cross Validation*, dengan nilai *K* yaitu 2, 5, dan 10. Setelah diketahui nilai masing – masing akurasi dari *K* 2, 5, & 10, selanjutnya dilakukan validasi dengan memilih akurasi di atas 50% dan menggunakan 14 *unseen* data test sebagai data uji pada masing – masing pengujian *K* Fold. Nilai akurasi dengan penurunan terkecil akan digunakan pada *website* sistem pendukung keputusan.

B. Perhitungan K Fold Cross Validation

Setelah data dibagi menjadi 50 data dan 14 *unseen* data, berikutnya yaitu melakukan pembagian data dengan menggunakan *K* 2, 5, & 10. Di bawah ini contoh perhitungan dari *K* 10 Fold Pengujian Ke – 5.

Tabel 11. Hasil Klasifikasi

	1-5	6-10	11-15	16-20	21-25	41-45	46-50
P1							
P2							
P3							
P4							
P5						
P6							
P7							
P8							
P9							
P10							

Pada tabel di atas pengujian ke 1 hingga ke 10 menggunakan 45 data latih (kolom tidak terarsir) dan 5 data uji (kolom terarsir), data tersebut diambil dari 50 data yang telah dibagi sebelumnya. Setelah dibagi menjadi data latih dan data uji, maka selanjutnya akan dilakukan perhitungan menggunakan metode gaussian naïve bayes. Berikut merupakan contoh dari pengujian ke 5.

Tabel 12. Data Latih & Uji 10 Fold Pengujian Ke - 5

No	Nama	Speed	Power	Stamina	Agility	Kedisiplinan	Gerak/Teknik	Kategori
1	Fat’Hur	77	75	71	70	67	66	Fight
2	Arif	60	69	61	68	69	59	Fight
3	Yaqin	70	67	64	63	61	60	Fight
4	ALFIN	72	69	66	65	63	62	Fight
5	Saputra	72	69	66	65	63	62	Fight
.....								
45	Fitri	61	69	72	61	63	76	TGR
No	Nama	Speed	Power	Stamina	Agility	Kedisiplinan	Gerak/Teknik	Kategori
1	Bandi	73	70	67	70	63	62	Fight
2	Mahmud	72	69	66	65	63	62	Fight
3	Rama	60	56	53	52	51	49	SerHin
4	Wahidah	64	61	58	56	55	54	TGR
5	Alfin B	77	74	71	69	68	66	TGR

Setelah 50 data dibagi masing – masing data latih 45 data dan data uji 5 data, maka selanjutnya mencari nilai probabilistik menggunakan rumus probabilistik dan nilai mean atau rata – rata kategori Fight atribut Speed menggunakan data latih dengan rumus mean.

Probabilitas	
Atribut	Nilai
Fight	0.333333333333333
TGR	0.333333333333333
Serang Hindar	0.333333333333333

Gambar 3. Tampilan Hasil Probabilitas Kelas Pada Website

Setelah mendapatkan hasil probabilitas selanjutnya mencari nilai mean menggunakan data latih.

Mean						
Atribut	Speed	Power	Stamina	Agility	Kedisiplinan	Gerak Teknik
Fight	71	69.4	66.1333333333333	66.8666666666667	63.2	61.8
TGR	67.4666666666667	69.8666666666667	68.4	61.8	61.5333333333333	66.7333333333333
Serang Hindar	68.8	66.7333333333333	64.6666666666667	65.6666666666667	65.3333333333333	64.8666666666667

Gambar 4. Tampilan Hasil Mean / Rata - Rata Pada Website

Selanjutnya menghitung standard deviasi setiap kategori dan atribut menggunakan rumus standard deviasi.

Standar Deviasi						
Atribut	Speed	Power	Stamina	Agility	Kedisiplinan	Gerak Teknik
Fight	4.3260011227791	2.472708174105	2.6149751088106	3.0441200151549	3.3636714634883	2.4260491103261
TGR	9.8115578103921	5.3833959896103	5.7171421432141	9.7189064640598	7.0291908360439	7.3917586025313
Serang Hindar	6.1783030494974	5.2435901913897	4.1346388107352	4.7908643220593	4.0118870990144	5.2354105674268

Gambar 5. Tampilan Hasil Standard Deviasi Pada Website

Berikutnya mencari hasil probabilistic pada data uji dari setiap atribut untuk masing – masing kelas, menggunakan rumus *gaussian naïve bayes*. Contoh perhitungan atribut *speed* kategori *fight* data uji ke – 1.

$$p(x) = \frac{1}{\sqrt{2} \times 3,14 \times 4,326} e^{-\frac{(73-71)^2}{2(4,326^2)}} = 0,1724$$

Keseluruhan data latih dihitung menggunakan cara di atas sehingga dapat memperoleh keseluruhan nilai probabilistik. Berikutnya mencari nilai klasifikasi dengan cara mengalikan hasil probabilistik dengan nilai probabilitas setiap kelas. Contoh perhitungan pada data uji ke -1 kategori *fight*.

Nilai di bawah ini dari kiri ke kanan adalah *Speed, Power, Stamina, Agility, Kedisiplinan, Gerak/Teknik* dan Probabilitas Kelas.

$$0.1724 \times 0.2436 \times 0.2335 \times 0.1346 \times 0.2171 \times 0.2553 \times 0,33 = 0,00002442$$

Hitung semua data berdasarkan nilai probabilistik lalu membandingkannya pada setiap kategori yang ada. Hasil terbesar dipilih sebagai hasil klasifikasi. Angka 1 kategori *fight*, 2 sebagai kategori *TGR* dan 3 sebagai kategori *serang hindar*.

Tabel 13. Hasil Klasifikasi Atlet

No	Nama	Klasifikasi	Prediksi	Real	Kriteria
1	Bandi	0,00002442	1	1	True
2	Mahmud	0,00004036	1	1	True
3	Rama	0,0000000002736	2	3	False
4	Wahidah	0,00000001965	2	2	True
5	Alfin B	0,0000007151	2	2	True

Kemudian hitung akurasi dari hasil klasifikasi di atas:

$$Akurasi = \frac{4}{5} \times 100\% = 80\%$$

Maka 80% merupakan hasil akurasi dari 10 Fold pengujian ke -5. Di bawah ini merupakan hasil akurasi dari 2, 5, dan 10 Fold.

Tabel 14. Hasil Akurasi K Fold Cross Validation K 2,5, dan 10

Fold	Skenario	Data Uji	Data Latih	Akurasi
2	1	25	25	68%
	2	25	25	68%
5	1	10	40	60%
	2	10	40	50%
	3	10	40	50%
	4	10	40	90%
	5	10	40	50%
10	1	5	45	60%
	2	5	45	80%
	3	5	45	60%
	4	5	45	80%
	5	5	45	80%
	6	5	45	40%
	7	5	45	80%
	8	5	45	80%
	9	5	45	40%
	10	5	45	20%

C. Penggunaan Unseen Data

Setelah keseluruhan akurasi telah di ketahui pada tabel 15, selanjutnya menggunakan *unseen* data pada data yang memiliki akurasi di atas 50%. Di bawah ini merupakan contoh hasil dari penggunaan *unseen* data.

Tabel 15. Hasil Akurasi Setelah Menggunakan Unseen Data

Fold	Pengujian Ke-	Data Uji	Data Latih	Akurasi Sebelumnya	Akurasi Terbaru	Penurunan Akurasi
5	4	14	40	90%	50%	40%

Terjadi penurunan akurasi sebanyak 40% dari akurasi sebelumnya yaitu 90%, hasil lengkap dari pengujian menggunakan data validasi dapat di lihat pada Gambar 6 di bawah ini:

Nama	Hasil	Prediksi	Akurasi
<ul style="list-style-type: none"> • Kia Dzaky E • Abdur Rosyid Bachtiar • Robiatul Andawiyah • Ahmad Ifan Hakiki • Meta Gadiecha W • Yuslianto • Dian Amaniatul Fitri • Anggi Sofy Anjeil • Ali Wajihah • Nur Fajar Eka Saputra • Moh Wahyu Al Waris • Syaifuddin • Satrio Budil • Faizah Aprilia Maulida 	<ul style="list-style-type: none"> • 3.47618964804E-5 • 4.0392771271857E-5 • 8.5714239055807E-7 • 2.4344732144505E-7 • 3.3514908856027E-8 • 2.1015069892522E-8 • 1.1819529495121E-5 • 1.92688107269E-7 • 1.1374683501533E-5 • 3.2058890181106E-6 • 8.0686718128235E-7 • 3.47618964804E-5 • 8.4454224768475E-7 • 1.649893262749E-8 	<ul style="list-style-type: none"> • Fight • Fight • Fight • TGR • TGR • TGR • Fight • TGR • Fight • TGR • Fight • Serang Hindar • TGR • Fight • Serang Hindar • TGR 	Akurasi = 57.142857142857 %

Gambar 6. Hasil Akurasi Unseen Data Pada Website

Pada Gambar 6 di atas merupakan hasil penggunaan *unseen* data pada website.

Tabel 16. Keseluruhan Akurasi Setelah Menggunakan *Unseen* Data

Fold	Pengujian Ke-	Data Uji	Data Latih	Akurasi Sebelumnya	Akurasi Terbaru	Penurunan Akurasi
2	1	14	25	68%	50%	18%
	2	14	25	68%	42,8%	25,1%
5	1	14	40	60%	42,8%	17,1%
	2	14	40	50%	-	-
	3	14	40	50%	-	-
	4	14	40	90%	50%	40%
	5	14	40	50%	-	-
10	1	14	45	60%	50%	10%
	2	14	45	80%	50%	30%
	3	14	45	60%	35,7%	24,2%
	4	14	45	80%	50%	30%
	5	14	45	80%	57,1%	22,8%
	6	14	45	40%	-	-
	7	14	45	80%	50%	30%
	8	14	45	80%	42,8%	37,1%
	9	14	45	40%	-	-
	10	14	45	20%	-	-

Pada Tabel 16 di atas terdapat beberapa data yang memiliki akurasi di atas 50%, yang terkecil terdapat pada Fold ke 5 percobaan ke – 1, Fold ke – 10 pengujian ke – 1, & 3, yang masing – masing memiliki akurasi sebesar 60%. Pada Fold ke – 2 pengujian ke -1 & 2 memiliki akurasi 68%. Dan akurasi sebesar 80% pada Fold ke – 10 pengujian ke – 2, 4, 5, 7, & 8. Dan akurasi tertinggi terdapat pada Fold ke – 5 pengujian ke – 4 yang memiliki akurasi 90%. Dari beberapa akurasi yang telah disebutkan beberapa akurasi mengalami penurunan akurasi yang tinggi, salah satunya terdapat pada Fold ke – 5 pengujian ke – 4, yang mana akurasi sebelumnya 90% turun menjadi 50% setelah menggunakan *unseen* data, terjadi penurunan sebanyak 40% yang mana ini merupakan penurunan terbesar dari beberapa data yang menggunakan *unseen* data. Dan penurunan terkecil terdapat pada Fold ke – 10 pengujian ke – 1, yang memiliki akurasi sebesar 60% turun menjadi 50% setelah menggunakan *unseen* data. Dari beberapa data di tabel 4.10 di atas data pada Fold ke – 10 pengujian ke – 5 dipilih menjadi data yang akan digunakan pada *website* dikarenakan akurasi sebelumnya relatif tinggi yaitu 80% dan penurunan akurasi relatif rendah yaitu 22,8% menjadi 57,1%.

Untuk mengetahui Recall dan Presisi dari setiap kategori menggunakan *Confusion Matrix*, berikut merupakan hasil dari Recall dan Presisi dari ketiga kategori.

Tabel 17. Hasil *Confusion Matrix* Kategori *Fight*

No	Kategori Perbandingan	Prediksi	Kriteria
1	1	1	TP
2	1	1	TP
3	1	1	TP
4	3	1	FP
5	2	1	FP
6	2	2	TN
7	2	2	TN
8	1	2	FN
9	2	2	TN
10	2	2	TN
11	1	3	FN
12	3	3	TN
13	2	3	FN
14	1	3	FN

Berdasarkan Tabel 17 di atas *Recall* dan Presisi dari kategori *Fight* adalah 42,8% untuk *Recall* dan 60% Presisi.

Tabel 18. Hasil *Confusion Matrix* Kategori *Fight*

No	Kategori Perbandingan	Prediksi	Kriteria
1	1	1	TN
2	1	1	TN
3	1	1	TN
4	3	1	FN
5	2	1	FN
6	2	2	TP
7	2	2	TP
8	1	2	FP
9	2	2	TP
10	2	2	TP
11	1	3	FN
12	3	3	TN
13	2	3	FN
14	1	3	FN

Berdasarkan Tabel 18 di atas *Recall* dan Presisi dari kategori *Fight* adalah 80% untuk *Recall* dan 44,4% untuk Presisi.

Tabel 19. Hasil *Confusion Matrix* Kategori *Fight*

No	Kategori Perbandingan	Prediksi	Kriteria
1	1	1	TN
2	1	1	TN
3	1	1	TN
4	3	1	FN
5	2	1	FN
6	2	2	TN
7	2	2	TN
8	1	2	FN
9	2	2	TN
10	2	2	TN
11	1	3	FP
12	3	3	TP
13	2	3	FP

No	Kategori Perbandingan	Prediksi	Kriteria
14	1	3	FP

Berdasarkan Tabel 19 di atas *Recall* dan Presisi dari kategori Fight adalah 25% untuk *Recall* dan 25% untuk Presisi.

5. KESIMPULAN

A. Kesimpulan

Berdasarkan hasil penelitian klasifikasi atlet pencak silat perisai diri menggunakan metode gaussian naïve bayes berbasis website ini, di dapatkan akurasi sebesar 57,1%. Tingkat Presisi dan Recall pada setiap kategori di dapatkan hasil sebagai berikut:

- 1) Kategori Fight: Recall 42,8%, Presisi 60%
- 2) Kategori TGR: Recall 80%, Presisi 44.4%
- 3) Kategori Serang Hindar: Recall 25%, Presisi 25%

B. Saran

Berdasarkan penelitian yang telah di lakukan maka saran peneliti untuk penelitian selanjutnya ialah menggunakan metode klasifikasi yang berbeda, dan menggunakan data lebih banyak.

6. DAFTAR PUSTAKA

- Bustami. (2014). Penerapan Algoritma Naive Bayes. *Jurnal Informatika*, 8(1), 884–898.
- Fauzi, A. (2016). Sistem Pendukung Keputusan Dalam Menentukan Supplier Infrastruktur IT (Studi Kasus: PT. Cipta Karya Komputer). *SWABUMI*, IV(2), 121–128.
- Hutahaean, W., & Hasiguan, P. S. (2021). Sistem Pendukung Keputusan Penentuan Penerima Bantuan Bedah Rumah Menggunakan Metode Weighted Product Pada Kecamatan Borbor Wasindo. *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, 4(1), 113–117.
- Maryono, O. (1998). *Pencak Silat Merentang Waktu*. Pustaka Pelajar.
- Pratama, Rendra, Y., & Trilaksana, A. (2018). Perkembangan Ikatan Pencak Silat Indonesia (IPSI) Tahun 1948-1973. *e-Journal Pendidikan Sejarah*, 6(3), 1–10.
- Rohani, A., Taki, M., & Abdollahpour, M. (2018). A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I). *Renewable Energy*, 115, 411–422.
- Rustam, Z., & Audia Ariantari, N. P. A. (2018). Support Vector Machines for Classifying Policyholders Satisfactorily in Automobile Insurance. *Journal of Physics: Conference Series*, 1028(1), 1–8.
- Santra, a. K., & Christy, C. J. (2012). Genetic Algorithm and Confusion Matrix for Document Clustering. *International Journal of Computer Science*, 9(1), 322–328.