

# KLASIFIKASI NAIVE BAYES KEPARAHAN TRAUMA PASIEN MENGUNAKAN DATA *NEURO COGNITIVE* DAN DATA *PHYSIOLOGIC* DENGAN PYTHON

(*Naive Bayes Classification of Patient's Trauma Severity Using Neuro Cognitive Data and Physiologic Data with Python*)

Puja Aditya Winata

Universitas Negeri Semarang, Sekaran Gunungpati, Semarang  
e-mail: adityawinata@students.unnes.ac.id

**Abstract.** Trauma is a mental health condition that can lead to post-traumatic stress disorder (PTSD). Several studies have shown that there are signs of PTSD in COVID survivors. This shows that the ongoing COVID-19 pandemic can increase PTSD patients. Therefore, it is important to diagnose the severity of trauma so that appropriate treatment can be carried out and to minimize the risk of increasing PTSD patients during the pandemic. The diagnosis of severity in this paper uses a machine learning classification method using the Naïve Bayes algorithm. The formation of this model is expected to help experts in correlating neuro-cognitive data and physiologic data to produce patient trauma diagnoses easily, quickly and precisely. The results of the accuracy using the confusion matrix obtained are 99.537%, so the formed model shows very good performance.

**Keywords:** Classification, Naïve Bayes, Post-traumatic stress disorder, Trauma

## 1. Pendahuluan

Penggunaan *machine learning* kini menjamur dalam berbagai bidang. Mulai dari bidang pendidikan, pertanian, medis hingga psikologi. Dwyer *et al.*, [1] menuliskan *machine learning* secara luas didefinisikan sebagai strategi komputasi yang secara otomatis belajar menentukan metode dan parameter untuk mencapai solusi permasalahan yang optimal, daripada diprogram oleh manusia untuk menyelesaikan solusi tetap, selain itu *machine learning* merupakan subbidang dari *artificial intelligence* dikarenakan proses belajar dalam *machine learning* mensimulasikan aspek kecerdasan manusia.

Trauma merupakan kondisi yang banyak dialami orang-orang pasca kecelakaan, bencana alam atau bahkan pandemi COVID-19. *American Psychological Association* (APA) menjelaskan bahwa trauma merupakan respon emosional terhadap peristiwa-peristiwa yang mengerikan seperti bencana alam, pemerkosaan, maupun kecelakaan [2]. Reaksi jangka panjang pada trauma yang dialami adalah emosi yang tidak terduga, kilas balik kejadian mengerikan, *strained relationship* (hilangnya kepercayaan antar hubungan), atau bahkan gejala fisik seperti sakit kepala atau mual. Trauma yang berkepanjangan dapat menyebabkan gangguan *post-traumatic stress disorder* (PTSD). Bisson *et al.*, [3] menuliskan bahwa PTSD adalah gangguan mental yang dapat berkembang setelah mengalami peristiwa yang mengancam atau mengerikan. Data dari Perhimpunan Dokter



Spesialis Kedokteran Jiwa Indonesia (PDSKJI), menunjukkan 84% dari 2113 orang mengalami trauma psikologis selama masa pandemi dari tahun 2020 sampai 2022 [4]. Kemudian penelitian yang dilakukan oleh Kariisa et al., [5] menunjukkan bahwa ditemukan gejala PTSD pada 30,2% pasien penyintas COVID-19. Spencer-Segal et al., [6] juga dalam artikelnya menuliskan bahwa pasien COVID yang dirawat di rumah sakit memiliki tingkat kecemasan, *post-traumatic stress*, dan kesepian yang tinggi dibandingkan dengan pasien bukan COVID yang dirawat di rumah sakit. Hal tersebut menunjukkan bahwa pandemi COVID-19 yang terjadi saat ini dapat meningkatkan jumlah pengidap PTSD, sehingga penting dilakukan pemeriksaan tingkat keparahan trauma agar dapat dilakukan pengobatan yang sesuai untuk menekan jumlah pengidap PTSD selama masa pandemi.

*Cognitive neuroscience* merupakan ilmu yang mempelajari pemikiran dan perilaku yang diinformasikan oleh penemuan-penemuan dari ilmu saraf tentang sifat fisik dari proses otak [7]. Seiring dengan perkembangan teknologi, penelitian *neuro cognitive* dilakukan dengan menggunakan alat *brain imaging*, seperti *Electroencephalograms* atau EEG. Dilansir dari Halodoc [8], EGG atau *electroencephalography* merupakan prosedur pemeriksaan dengan menggunakan alat yang berfungsi untuk mempelajari rekaman aktivitas listrik yang terjadi di otak. EEG dilakukan untuk menegakkan diagnosis. Menurut Kamus Besar Bahasa Indonesia (KBBI), fisiologi merupakan cabang dari ilmu biologi yang mempelajari fungsi dan kegiatan kehidupan atau zat hidup meliputi organ, jaringan, atau sel [9]. Schnädelbach [10] menuliskan data fisiologi mengacu pada data seperti detak jantung, pernapasan, dan aktivitas otak yang dapat diukur, dimanipulasi dan sampai batas tertentu ditafsirkan untuk memahami keadaan emosi.

*Data mining* merupakan ekstraksi pola atau pengetahuan yang menarik dari sejumlah data yang sangat besar [11]. Salah satu teknik yang digunakan dalam data mining adalah klasifikasi. Klasifikasi termasuk dalam *supervised learning* yang banyak digunakan untuk proses prediksi. *Supervised learning* sendiri merupakan teknik menggunakan kumpulan data yang memiliki label output, dan klasifikasi melakukan prediksi untuk kasus mendatang berdasarkan kumpulan data sebelumnya [12]. Algoritma yang dikenal untuk melakukan proses klasifikasi salah satunya adalah Naïve Bayes. Algoritma Naïve Bayes juga banyak digunakan untuk melakukan klasifikasi pada gangguan mental. Penelitian terdahulu yang menggunakan algoritma Naïve Bayes untuk melakukan klasifikasi gangguan mental adalah penelitian yang dilakukan oleh Dewi [13] yaitu klasifikasi ASD atau *autism spectrum disorders* pada anak menggunakan data DSM-5 dan menghasilkan nilai akurasi tertinggi sebesar 98,6301% dengan rasio data latih dan uji 1:1. Selain itu penelitian yang dilakukan oleh Sugari et al., [14] tentang perbandingan algoritma C4.5 dan Naïve Bayes untuk deteksi gangguan autisme pada anak menunjukkan bahwa algoritma Naïve Bayes memiliki nilai akurasi lebih tinggi, yakni sebesar 73,33% dibandingkan algoritma C4.5 dengan nilai akurasi 72%.

Berangkat dari trauma yang saat ini kian meningkat terlebih bagi pasien penyintas COVID selama pandemi, dalam artikel ini penulis melakukan klasifikasi algoritma Naïve Bayes untuk mendeteksi tingkat keparahan trauma menggunakan data *neuro cognitive* dan data *physiologic* pasien, sehingga penelitian ini diharapkan dapat membantu para pakar dalam mengkorelasikan data *neuro cognitive* dan data *physiologic* untuk menghasilkan diagnosis trauma pasien dengan mudah, cepat dan tepat.

## 2. Metodologi

### 2.1 Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu algoritma yang dipakai untuk proses klasifikasi *machine learning*. Sumathi & Esakkirajan [15] dalam bukunya menuliskan bahwa Naïve Bayes merupakan algoritma yang didasarkan pada Teorema Bayes di pertengahan abad ke delapan belas. Teorema Bayes didefinisikan sebagai berikut.

$$P(B) = \frac{P(A)P(A)}{P(B)} \quad (1)$$

Dimana  $A$  dan  $B$  merupakan kejadian, dan  $P(B) \neq 0$ .  $P(A)$  merupakan peluang kejadian  $A$  atau *prior probability*,  $P(B)$  merupakan peluang kejadian  $B$ ,  $P(A|B)$  merupakan peluang terjadi kejadian  $A$  dan  $B$  telah terjadi atau disebut *posteriori probability*, dan  $P(B|A)$  merupakan peluang terjadi kejadian  $B$  dan  $A$  telah terjadi.

### 2.2 Data

Dataset yang digunakan dalam artikel ini diambil dari Kaggle, yaitu *Multiple data sources analysis of Trauma Patients* [16]. Dataset terdiri dari 3599 data dan 14 *features*. 5 *feature* merupakan data *neuro cognitive*, 7 *feature* merupakan data fisiologis, 1 *feature* merupakan data ID pasien, dan 1 *feature* target. *Features* yang ada didalam data dapat dilihat pada Tabel 1.

```
['Hair_Phenotype',  
 'heart_rate',  
 'skin_conductance',  
 'skin_temperature',  
 'cortisol_level',  
 'Systolic_BP',  
 'Diastolic_BP',  
 'Trauma_Severity']
```

Gambar 1. Data Kategorial

Tabel 1. *Feature* dalam dataset

<b>Data ID Pasien</b>
Patinet_ID
<b>Data Neuro Cognitive</b>
EEG_Signal_Amplitude
EEG_Delta_band
EEG_Theta_band
EEG_Alpha_band
EEG_Beta_band
<b>Data Fisiologi</b>
Hair_Phenotype
Heart_rate
Skin_conductance
Skin_temperature
Cortisol_level
Systolic_BP
Diastolic_BP
<b>Feature Target</b>
Trauma_severity

Berdasarkan output Gambar 1, ada 8 data kategorial, kedelapan data tersebut dapat dilihat dalam Gambar 1.

Masing-masing data kategorial memiliki label tersendiri. Untuk label dari masing-masing data kategorial dapat dilihat pada Tabel 2, sedangkan tampilan 10 data teratas dapat dilihat pada Gambar 2.

### 2.3 Python

Python merupakan salah satu bahasa pemrograman yang telah dikenal oleh banyak orang. Shukla & Shukla [17] menjelaskan bahwa faktor penting dalam utilitas Python sebagai bahasa komputasi untuk ilmuwan dan insinyur adalah sintaks yang jelas. Sintaks yang jelas ini dapat membuat kode mudah dimengerti dan mudah untuk *maintenance*. Beberapa spesifikasi sintaks ini termasuk blok kode yang ditentukan oleh *indentation*, ekstensif ruang nama modul, konstruksi pengulangan yang mudah dibaca, penanganan pengecualian, dan dokumentasi string.

### 2.4 Confusion Matrix

Sumathi & Esakkirajan [14] kembali menuliskan *confusion matrix* merupakan tabel  $N \times N$  yang merangkum keakuratan model prediksi klasifikasi. Dimana  $N$  mewakili jumlah kelas. Dalam masalah klasifikasi biner,  $N = 2$ . Secara sederhana *confusion matrix* adalah korelasi antara label aktual dan label prediksi model. Dalam artikel ini, *confusion matrix* dipakai untuk mengukur performa model klasifikasi Naïve Bayes yang telah dibentuk. Rumus menghitung nilai akurasi menggunakan *confusion matrix* dapat dicari

menggunakan rumus (2) berikut.

$$Accuracy = \frac{TP+TN}{(TP+TN+FN+FP)} \quad (2)$$

dimana TP merupakan *true positive*, TN merupakan *true negative*, FN merupakan *false negative*, dan FP merupakan *false positive*.

Tabel 2. Label dari masing-masing *features* kategorial

Hair_ Phenotype	Heart_ rate	skin_ conductance	cortisol_ level	Systolic_ BP	Diastolic_ BP	Taruma_ severity
Curly_hair	Medium_ PulseRate	Normal_ Conductance	Average CL	Range2_ LowSyst olic	NormalDi systolic	Medium _Severity
Wavy_Hair	High_ PulseRate	Low_ Conductance	Below_ Average CL	Range3_ LowSyst olic	LowDiSy stolic	Low_Sev erity
Straight_ Hair	Low_ PulseRate	High_ Conductance	Above_ Average CL	Range1_ LowSyst olic	verylowD iSystolic	High_Se verity
No_Hair						

	Patinet_ID	EEG_Signal_Amplitude	EEG_Delta_band	EEG_Theta_band	EEG_Alpha_band	EEG_Beta_band	Hair_Phenotype	heart_rate	skin_conductance	
0	180203		56	2.4786	5.5748	11.7319	23.9909	Curly_hair	Medium_PulseRate	Normal_Conductance
1	152268		97	3.2531	6.4658	13.1411	28.3962	Curly_hair	High_PulseRate	Low_Conductance
2	157399		83	3.6325	6.0053	13.6766	26.0487	Wavy_hair	High_PulseRate	Low_Conductance
3	131849		58	2.9477	5.5462	10.3739	22.0865	Straight_hair	Medium_PulseRate	Normal_Conductance
4	164593		22	1.9366	4.3574	8.9079	18.7077	Curly_hair	Low_PulseRate	High_Conductance
5	173369		62	3.9013	6.2357	12.5692	27.3996	Curly_hair	High_PulseRate	Low_Conductance
6	115630		76	3.6032	6.8745	15.8141	27.0002	No_hair	High_PulseRate	Low_Conductance
7	142490		39	2.5990	5.1169	11.6822	20.3183	No_hair	Medium_PulseRate	Normal_Conductance
8	130361		38	2.3969	5.5707	10.7618	23.1681	Straight_hair	Medium_PulseRate	Normal_Conductance
9	128397		97	3.5554	6.8908	13.7521	26.2053	Straight_hair	High_PulseRate	Low_Conductance

	Patinet_ID	skin_temperature	cortisol_level	Systolic_BP	Diastolic_BP	Trauma_Severity
0	180203	Normal_Temperature	AverageCL	Range2_LowSystolic	NormalDiSystolic	Medium_Severity
1	152268	Fever	Below_AverageCL	Range3_LowSystolic	LowDiSystolic	Low_Severity
2	157399	Fever	Below_AverageCL	Range3_LowSystolic	LowDiSystolic	Low_Severity
3	131849	Normal_Temperature	AverageCL	Range2_LowSystolic	NormalDiSystolic	Medium_Severity
4	164593	Low_Temperature	Above_AverageCL	Range1_LowSystolic	VerylowDiSystolic	High_Severity
5	173369	Fever	Below_AverageCL	Range3_LowSystolic	LowDiSystolic	Low_Severity
6	115630	Fever	Below_AverageCL	Range3_LowSystolic	LowDiSystolic	Low_Severity
7	142490	Normal_Temperature	AverageCL	Range2_LowSystolic	NormalDiSystolic	Medium_Severity
8	130361	Normal_Temperature	AverageCL	Range2_LowSystolic	NormalDiSystolic	Medium_Severity
9	128397	Fever	Below_AverageCL	Range3_LowSystolic	LowDiSystolic	Low_Severity

Gambar 2. Tampilan 10 data teratas

### 3. Hasil dan Pembahasan

#### 3.1 Data Preprocessing dan Proses Training

Data yang digunakan terlebih dahulu harus melalui proses *preprocessing*. Tahapan *preprocessing* data ini bertujuan untuk menyiapkan data mentah sebelum masuk ke dalam tahap pengolahan data lebih lanjut. Dalam tahap ini dilakukan pemilihan *feature* yang

akan dipakai, pemeriksaan *missing value* atau data yang bernilai *null*, menghilangkan duplikasi data, pembagian dataset menjadi data *train* dan data *testing*, melabeli atau *encoding* data kategorial, serta *feature scaling*.

```
Index(['Patinet_ID', 'EEG_Signal_Amplitude', 'EEG_Delta_band',
      'EEG_Theta_band', 'EEG_Alpha_band', 'EEG_Beta_band', 'Hair_Phenotype',
      'heart_rate', 'skin_conductance', 'skin_temperature', 'cortisol_level',
      'Systolic_BP', 'Diastolic_BP', 'Trauma_Severity'],
      dtype='object')
```

Gambar 3. Feature dalam dataset

Berdasarkan *feature* pada Gambar 3 diatas, Patient\_ID akan dihapus karena tidak akan digunakan dalam proses pengolahan data menggunakan Naïve Bayes.

```
df.isnull().sum()
Patinet_ID      0
EEG_Signal_Amplitude  0
EEG_Delta_band  0
EEG_Theta_band  0
EEG_Alpha_band  0
EEG_Beta_band   0
Hair_Phenotype  0
heart_rate      0
skin_conductance  0
skin_temperature  0
cortisol_level  0
Systolic_BP     0
Diastolic_BP    0
Trauma_Severity  0
dtype: int64

df.duplicated().any()
False
```

Gambar 4. Output *missing value*

Setelah melakukan pemeriksaan *missing value* dan duplikasi data, Gambar 4 menunjukkan tidak ditemukannya *missing value* dan duplikasi data. Pembagian data *train* dan data *testing* adalah 70% data digunakan untuk data *train* dan 30% data digunakan untuk data *testing*. Pembagian data lebih lanjut dapat dilihat pada Tabel 3.

Tabel 3. Pembagian Data

Data Train	Data Testing
2519	1080

Hair_Phenotype	heart_rate	skin_conductance	skin_temperature	cortisol_level	Systolic_BP	Diastolic_BP
0	2	2	2	1	1	1
0	0	1	0	2	2	0
3	0	1	0	2	2	0
2	2	2	2	1	1	1
0	1	0	1	0	0	2

Gambar 5. *Feature* kategorial setelah proses encoding

Gambar 5 menunjukkan data kategorial yang telah dilakukan proses *encoding* atau mentranformasi data ketegorial ke dalam bentuk numerik.

EEG_Signal_Amplitude	EEG_Delta_band	EEG_Theta_band	EEG_Alpha_band	EEG_Beta_band	Hair_Phenotype	heart_rate	skin_conductance	skin_temperature
1.044444	0.904668	0.446203	0.934568	0.947554	0.000000	-0.5	0.0	-0.5
-0.666667	-0.775488	-0.548399	-0.396295	-0.542917	1.333333	0.0	-0.5	0.0
-0.066667	-0.000614	0.191987	-0.037521	0.144080	0.666667	0.5	0.5	0.5
0.533333	0.354374	0.475440	1.302368	0.435799	-0.666667	-0.5	0.0	-0.5
1.155556	0.993108	0.346890	0.370669	0.694666	-0.666667	-0.5	0.0	-0.5

Gambar 6. *Feature scalling*

Gambar 6 menunjukkan data setelah dilakukan proses *feature scalling*. Proses ini berguna untuk menyamakan rentan nilai (*scale*) pada data. Algoritma Naïve Bayes yang digunakan untuk proses *training* adalah *Gaussian Naive Bayes classifier*, sehingga perlu dilakukan pengecekan distribusi normal pada dataset.

Dengan menggunakan hipotesis berikut:

$H_0$  = Variable berdistribusi normal (p-value  $\geq 0.05$ )

$H_1$  = Variable tidak berdistribusi normal (p-value  $< 0.05$ )

Dan pengujian normalitas menggunakan fungsi *multivariate\_normality*, diperoleh hasil seperti berikut

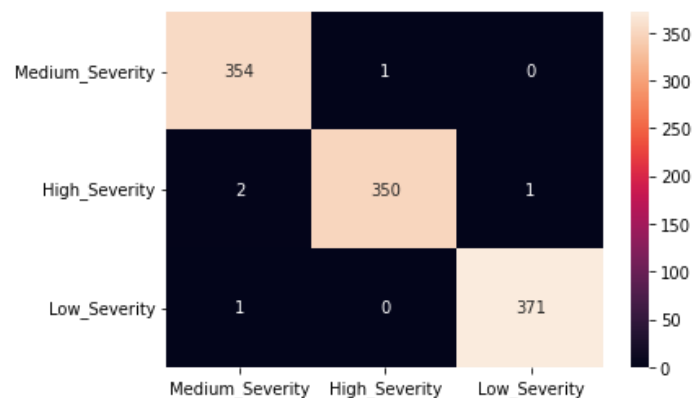
```
HZResults(hz=1.000882614823801, pval=0.1173960548015302, normal=True)
```

Diperoleh p-value = 0.117 dan p-value  $\geq 0.05$  maka  $H_0$  diterima.

Setelah melakukan *preprocessing data*, dilanjutkan dengan melakukan proses *training*.

### 3.2 Proses Testing

Untuk mengukur performa model klasifikasi Naïve Bayes yang telah dibentuk melalui proses *training*, dilakukan proses *testing* atau pengujian pada model dengan menggunakan data *testing*. Pada artikel ini, *confusion matrix* digunakan untuk mengukur tingkat akurasi model klasifikasi Naïve Bayes yang telah dilatih. Setelah dilakukan proses *testing* diperoleh *confusion matrix* yang ditampilkan pada Gambar 7 berikut.



Gambar 7. *Confusion matrix* model klasifikasi yang telah dibentuk

Berdasarkan Gambar 7 diatas, maka diperoleh hasil sebagai berikut.

1. Prediksi keparahan medium dan nilai aktualnya juga medium sebanyak 354
2. Prediksi keparahan tinggi tetapi nilai aktualnya medium sebanyak 2
3. Prediksi keparahan rendah tetapi nilai aktualnya medium sebanyak 1
4. Prediksi keparahan medium tetapi nilai aktual tinggi sebanyak 1
5. Prediksi keparahan tinggi dan nilai aktualnya juga tinggi sebanyak 350
6. Prediksi keparahan rendah tetapi nilai aktualnya tinggi sebanyak 0
7. Prediksi keparahan medium, tetapi nilai aktualnya rendah sebanyak 0
8. Prediksi keparahan tinggi, tetapi nilai aktualnya rendah sebanyak 1
9. Prediksi keparahan rendah, dan nilai aktualnya juga rendah 371

Nilai akurasi berdasarkan rumus (2) diatas, maka diperoleh.

$$Accuracy = \frac{PrediksiBenar_{medium} + PrediksiBenar_{tinggi} + PrediksiBenar_{rendah}}{Total Prediksi}$$

$$= \frac{(354 + 350 + 371)}{(354 + 350 + 371 + 1 + 1 + 1 + 2)} = \frac{1075}{1080} = 0,9953 = 99,537\%$$

Selain nilai akurasi *testing*, diperlukan juga nilai akurasi *training*, hal ini bertujuan untuk mengetahui apakah terjadi *overfitting* dan *underfitting*.

Training set score: 0.9933  
Test set score: 0.9954

Gambar 8. Hasil akurasi *training* dan *testing*

Gambar 8 menunjukkan nilai akurasi *training* sebesar 0.9933 sedangkan nilai akurasi *testing* sebesar 0.9954. Hal ini menunjukkan bahwa tidak terjadi *overfitting* dan *underfitting* dari model yang telah dibentuk.

#### 4. Kesimpulan

Berdasarkan hasil pengujian model klasifikasi Naïve Bayes dalam menentukan tingkat keparahan trauma menggunakan data *neuro cognitive* dan fisiologis (*physiologic*) pengukuran perfoma model dengan *confusion matrix* diperoleh nilai akurasi model sebesar 0,9953 atau sebesar 99,53%, nilai ini diperoleh dengan membagi data 70% untuk data *train* dan 30% untuk data *testing*. Selain itu, akurasi untuk proses *training* menunjukkan nilai 0,9933 atau sebesar 99,33%, hal tersebut menunjukkan tidak terjadi *overfitting* dan *underfitting*, sehingga dapat disimpulkan model yang telah dibangun mampu memprediksi tingkat keparahan menggunakan data *neuro cognitive* dan data fisiologis dengan sangat baik.



## Daftar Pustaka

- [1] Dwyer, D. B., Falkai, P., Koutsouleris, N., (2018), Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, **14**, 91–118. <https://doi.org/10.1146/annurev-clinpsy-032816-045037>
- [2] American Psychological Association. (n.d.). Trauma. Diakses pada 15 Juni 2022. <https://www.apa.org/topics/trauma#:~:text=Trauma%20is%20an%20emotional%20response,symptoms%20like%20headaches%20or%20nausea>.
- [3] Bisson, J. I., Cosgrove, S., Lewis, C., Roberts, N. P., (2015), Post-traumatic stress disorder. *BMJ* (Online), **351**. <https://doi.org/10.1136/bmj.h6161>
- [4] Perhimpunan Dokter Spesialis Kedokteran Jiwa Indonesia (PDSKJI), (2022), *Masalah Psikologis 2 Tahun Pandemi COVID-19 di Indonesia*. Diakses pada 15 Juni 2022. <http://pdsjki.org/home>
- [5] Kariisa, M., Scholl, L., Wilson, N., Seth, P., Hoots, B., (2019), Drug Overdose Deaths Involving Cocaine and Psychostimulants with Abuse Potential — United States, 2003–2017. *MMWR. Morbidity and Mortality Weekly Report*, **68(17)**, 388–395. <https://doi.org/10.15585/mmwr.mm6817a3>
- [6] Spencer-Segal, J. L., Smith, C. A., Slavin, A., Sampang, L., DiGiovine, D., Spencer, A. E., Zhang, Q., Horowitz, J., Vaughn, V. M., (2021), Mental health outcomes after hospitalization with or without COVID-19. *General Hospital Psychiatry*, **72**, 152–153. <https://doi.org/10.1016/j.genhosppsych.2021.07.004>
- [7] Goodenough, O. R. & Tucker, M., (2010), Law and cognitive neuroscience. *Annual Review of Law and Social Science*, **6**, 61–92. <https://doi.org/10.1146/annurev.lawsocsci.093008.131523>
- [8] Halodoc. (n.d.), *Pemeriksaan Electroencephalography (EEG)*, Diakses pada 15 Juni 2022. <https://www.halodoc.com/kesehatan/pemeriksaan-electroencephalography-eeeg>.
- [9] fisiologi. 2016. Pada KBBI Daring. Diambil 15 Juni 2022.
- [10] Schnädelbach, H., (2011), Physiological Data in Adaptive Architecture, *Adaptive Architecture Conference*, March, 1–15.
- [11] Lakshmi, B. N. & Raghunandhan, G. H., (2011), A conceptual overview of data mining. *Proceedings of National Conference on Innovations in Emerging Technology, NCOIET'11*, 27–32. <https://doi.org/10.1109/NCOIET.2011.5738828>
- [12] Al-Janabi, M. I., Qutqut, M. H., Hijjawi, M., (2018), Machine Learning Classification

Techniques for Heart Disease Prediction: A Review. *International Journal of Engineering & Technology*, **7(4)**, 5373–5379.  
<https://doi.org/10.14419/ijet.v7i4.28646>

- [13] Dewi, E. S., (2021), Klasifikasi Autism Spectrum Disorder Menggunakan Algoritma Naïve Bayes. *MATHunesa: Jurnal Ilmiah Matematika*, **9(1)**, 27–35.  
<https://doi.org/10.26740/mathunesa.v9n1.p27-35>
- [14] Sugari, B., Adidarma, D., & Budilaksono, S., (2019), Perbandingan Akurasi Algoritma C4.5 dan Naïve Bayes untuk Deteksi Dini Gangguan Autisme pada Anak. *Jurnal IKRA-ITH Informatika*, **3(1)**, 119–128.
- [15] Sumathi, S., & Esakkirajan, S., (2007), Data mining and data warehousing. *In Studies in Computational Intelligence*, **47**. [https://doi.org/10.1007/978-3-540-48399-1\\_10](https://doi.org/10.1007/978-3-540-48399-1_10)
- [16] Spot, It., (2022), *Multiple data sources analysis of Trauma Patients*. Kaggle. Diambil 15 Juni 2022. <https://www.kaggle.com/datasets/itspot/multiple-data-sources-analysis-of-trauma-patients>.
- [17] Shukla, S. S., & Shukla, A. (1985). On scientific computing. *Intelligent Instruments and Computers, Applications in the Laboratory*, **3(4)**, 19–20.