

Analisis Algoritma *Gaussian Naïve Bayes* Terhadap Klasifikasi Data Pasien Demam Tifoid (*Typhoid Fever*) Di Puskesmas Balung Jember

Dimas Widia Adi Nugroho^{1*}, Hardian Oktavianto², Dewi Lusiana³

Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember¹²

Sistem Informasi, Fakultas Teknik, Universitas Muhammadiyah Jember³

Email: dimaswanugroho@gmail.com^{1*}; hardian@unmuhjember.ac.id²; dewilusiana@unmuhjember.ac.id³

ABSTRAK

Penelitian ini bertujuan untuk menganalisis algoritma *Gaussian Naïve Bayes* dalam klasifikasi data pasien demam Tifoid di Puskesmas Balung Jember. Data yang digunakan terdiri dari 515 pasien, dengan partisi data menggunakan 20% sebagai data uji dan 80% sebagai data latih. Pemodelan data dilakukan dengan skenario K Fold Cross Validation menggunakan nilai $k = 2, 4, 5, 8, \text{ dan } 10$. Untuk mengatasi ketidakseimbangan data digunakan metode *Synthetic Minority Over Sampling Technique* (SMOTE) sehingga menjadi 337 data. Hasil pemodelan menggunakan skenario K Fold Cross Validation menunjukkan bahwa nilai rata-rata akurasi tertinggi sebesar 89,61%, nilai rata-rata presisi tertinggi sebesar 90,07%, dan nilai rata-rata recall tertinggi sebesar 89,62%. Pengujian model *Gaussian Naïve Bayes* yang dibangun menghasilkan akurasi sebesar 94%, presisi sebesar 90%, dan recall sebesar 93%. Pengujian juga dibandingkan dengan metode *Baseline*, yaitu *Random Guess* dan *Threshold*. Metode *Gaussian Naïve Bayes* menunjukkan hasil yang lebih unggul dibandingkan dengan *Random Guess*, dengan akurasi pemodelan sebesar 49,85% dan pada pengujian sebesar 37,86%. Selain itu, model *Gaussian Naïve Bayes* juga unggul dibandingkan dengan metode *threshold*, dengan akurasi pemodelan sebesar 73,3% dan pada pengujian sebesar 75,73%. Penelitian ini menunjukkan bahwa algoritma *Gaussian Naïve Bayes* efektif dalam klasifikasi data pasien demam Tifoid. Penggunaan SMOTE sebagai metode penyeimbang data juga berhasil meningkatkan performa model. Penelitian ini memberikan kontribusi dalam pemahaman dan penerapan algoritma klasifikasi untuk analisis data pasien demam tifoid.

Kata Kunci: *Gaussian Naïve Bayes*, Klasifikasi, Demam Tifoid

ABSTRACT

This study aims to analyze the *Gaussian Naïve Bayes* algorithm in classifying data of typhoid fever patients at Balung Jember Public Health Center. The data used consists of 515 patients, with a data partitioning of 20% for testing and 80% for training. Data modeling was performed using the K-Fold Cross Validation scenario with k values of 2, 4, 5, 8, and 10. To address the data imbalance, the *Synthetic Minority Over-sampling Technique* (SMOTE) method was employed, resulting in 337 data points. The results of the modeling using the K-Fold Cross Validation scenario showed that the highest average accuracy value was 89.61%, the highest average precision value was 90.07%, and the highest average recall value was 89.62%. The testing of the constructed *Gaussian Naïve Bayes* model yielded an accuracy of 94%, a precision of 90%, and a recall of 93%. The testing was also compared to *Baseline* methods, namely *Random Guess* and *Threshold*. The *Gaussian Naïve Bayes* method demonstrated superior results compared to *Random Guess*, with a modeling accuracy of 49.85% and a testing accuracy of 37.86%. Additionally, the *Gaussian Naïve Bayes* model outperformed the *Threshold* method, with a modeling accuracy of 73.3% and a testing accuracy of 75.73%. This research indicates that the *Gaussian Naïve Bayes* algorithm is effective in classifying data of typhoid fever patients. The use of SMOTE as a data balancing method also successfully improved the model's performance. This study contributes to the understanding and application of classification algorithms for the analysis of typhoid fever patient data.

Keywords: *Gaussian Naïve Bayes*, Classification, Typhoid Fever

1. PENDAHULUAN

Tifus adalah infeksi akut yang disebabkan oleh bakteri *Salmonella enterica*, secara ilmiah dikenal sebagai *Salmonella Typhi* (S.Typhi). Kasus seluruh dunia pada kasus ini diperkirakan mencapai 21 juta dengan 128.000 hingga 161.000 kematian per tahun, sebagian besar tersebar pada daerah Asia Selatan dan Tenggara. Di Indonesia, demam tifoid menjadi penyakit endemik yang perlu diwaspadai, terutama dalam menghadapi peningkatan kasus dan resistensi terhadap obat. Pencegahan dapat dilakukan dengan perubahan perilaku dan kebersihan, seperti mencuci tangan secara rutin dan menghindari kontak dengan orang sakit. Penelitian sebelumnya telah menunjukkan keberhasilan Algoritma *Gaussian Naïve Bayes* dalam klasifikasi penyakit seperti tifus, hepatitis C, dan kanker

dengan tingkat akurasi yang tinggi. Oleh karena itu, penelitian ini bertujuan untuk menganalisis penggunaan algoritma ini dalam klasifikasi data pasien demam tifoid. Tujuan pada penelitian ini adalah untuk mengetahui nilai akurasi, presisi yang diperoleh metode *Gaussian Naïve Bayes* dalam mengklasifikasi data pasien penderita penyakit tifus (tifoid) di Puskesmas Balung Jember. Manfaat penelitian ini adalah sebagai alternatif terhadap klasifikasi data pasien penderita tifus (tifoid) di Puskesmas Balung Jember dengan menggunakan algoritma *Gaussian Naïve Bayes* dan dapat digunakan untuk referensi dalam melakukan penelitian berikutnya.

2. KAJIAN PUSTAKA

A. Demam Tifoid

Demam tifoid, juga dikenal sebagai tifus, disebabkan oleh infeksi *Salmonella enterica serovar typhi* (S.Typhi), sedangkan infeksi oleh *Salmonella enterica serovar paratyphi* A, B, dan C menyebabkan *paratifus*. Kedua kondisi ini termasuk dalam demam enterik, dimana sekitar 90% kasus demam enterik merupakan tifus di daerah endemik (Linson dkk., 1981). Gejala klinis demam tifoid, sebuah penyakit sistemik akut, bervariasi mulai dari demam ringan, kelemahan, dan batuk ringan hingga gejala yang lebih parah seperti gangguan saluran pencernaan dan komplikasi lainnya.

Demam tifoid disebabkan oleh bakteri *Salmonella typhi* dan *Salmonella paratyphi*, yang termasuk dalam famili *Enterobacteriaceae* (Bhandari & Thada, 2022). Bakteri *Salmonella* berasal dari sumber seperti unggas, telur, dan penyu (Gu dkk., 2020). Manusia adalah satu-satunya reservoir *Salmonella Typhi*, dan infeksi dapat terjadi melalui konsumsi makanan atau minuman yang terkontaminasi dengan feses manusia yang mengandung bakteri tersebut. Kontaminasi makanan dengan feses atau dari lingkungan sekitarnya juga dapat menjadi sumber infeksi. Penggunaan antibiotik spektrum luas, seperti *streptomisin*, dan pola makan yang buruk dapat meningkatkan risiko terkena demam tifoid. Penggunaan antibiotik tersebut dapat merusak flora normal di usus yang berfungsi sebagai pertahanan terhadap infeksi (Bhandari & Thada, 2022).

B. Data Mining

Data mining adalah metode yang melibatkan penggalian informasi penting dari sejumlah besar data. Tujuan dari *data mining* adalah menemukan nilai-nilai informasi yang sebelumnya tidak diketahui secara manual dari *dataset* yang ada (Bustami, 2014). *Data mining* bertujuan untuk mempelajari apa yang terjadi di masa lalu dan menerapkannya di masa depan guna mendapatkan hasil yang lebih baik. Di bidang analisis data medis dan pembangunan model prediksi, *data mining* dikenal dengan berbagai metode dan solusi teknisnya. Proses *Knowledge Discovery from Data* (KDD) merupakan langkah penting dalam menemukan pengetahuan dari data. KDD adalah proses ekstraksi informasi berguna dan sebelumnya tidak diketahui dari kumpulan data (Hardoon dkk., 2007). Proses dalam *Knowledge Discovery from Data* (KDD) terdiri dari beberapa tahap sebagai berikut (Han dkk., 2012).

1. Pembersihan data (*data cleaning*): menghilangkan *noise* dan data yang tidak konsisten, serta menghapus data yang tidak lengkap pada tahap ini.
2. Integrasi data (*data integration*): menggabungkan beberapa sumber data menjadi penyimpanan data yang koheren.
3. Seleksi data (*data selection*): mengumpulkan data yang relevan dengan tugas analisis dari *database*, dengan mengurangi atribut dan melakukan kompresi data.
4. Transformasi data (*data transformation*): mengubah dan mengkonsolidasikan data ke dalam bentuk yang sesuai untuk proses penambangan (*mining*), melalui ringkasan atau penggabungan operasi.
5. Penambangan data (*data mining*): proses pencarian pola dan informasi menarik pada data terpilih menggunakan teknik tertentu, yang sangat tergantung pada tujuan dan proses KDD keseluruhan.
6. Evaluasi pola (*pattern evaluation*): mengidentifikasi pola yang relevan dan mendasar, serta mengevaluasi pengetahuan yang ditemukan.

7. Representasi pengetahuan (*knowledge presentation*): menyajikan pengetahuan yang ditemukan secara visual kepada pengguna untuk memahami hasil *data mining*, dan memberikan informasi tentang metode yang digunakan untuk memperoleh pengetahuan tersebut.

C. Klasifikasi

Klasifikasi adalah proses menemukan karakteristik data dan menugaskannya ke kelas yang sesuai berdasarkan karakteristik yang relevan. Pada proses ini, dilakukan pencarian untuk mencari properti objek, dan objek dengan properti yang sama dikelompokkan ke dalam kategori yang telah ditentukan sebelumnya. Proses klasifikasi melibatkan perbandingan antara data yang telah ada sebelumnya, yang disebut data pelatihan, dengan data baru atau data uji. Hal ini memberikan kesempatan untuk menguji data tersebut. Untuk melakukan klasifikasi, diperlukan catatan yang memiliki atribut target atau judul. Tujuan dari klasifikasi adalah memprediksi kelas objek dalam setiap data yang diberikan. Tugas klasifikasi umumnya dimulai dengan kumpulan data yang memiliki kelas yang sudah diketahui. Klasifikasi biner adalah bentuk paling sederhana dari masalah klasifikasi. Ada beberapa algoritma yang dapat digunakan dalam proses klasifikasi, seperti *Decision Tree C4.5*, *Naïve Bayes*, dan *K-Nearest Neighbor (KNN)*.

D. Gaussian Naïve Bayes

Naïve Bayes adalah metode klasifikasi yang menggunakan pendekatan probabilistik dan statistik yang dikembangkan oleh Thomas Bayes, seorang ilmuwan Inggris. Metode ini memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu. "Naïve" dalam *Naïve Bayes* mengacu pada asumsi bahwa variabel-variabel yang terlibat dalam klasifikasi saling independen secara kondisional. Dalam klasifikasi *Naïve Bayes*, diasumsikan bahwa keberadaan atau tidaknya suatu fitur dalam satu kelas tidak memiliki hubungan dengan fitur-fitur di kelas lain. *Naïve Bayes* dapat diperluas untuk atribut dengan nilai riil dengan mengadopsi distribusi *Gaussian*. Metode *Gaussian Naïve Bayes* digunakan untuk memodelkan distribusi data dengan menggunakan nilai rata-rata (*mean*) dan standar deviasi dari data pelatihan. Dalam metode ini, dihitung probabilitas nilai input untuk setiap kelas. Dengan atribut bernilai riil, kita dapat menghitung rata-rata dan standar deviasi dari nilai input (x) untuk setiap kelas distribusi. Selain menghitung probabilitas untuk setiap kelas, kita juga perlu menyimpan nilai rata-rata (*mean*) dan standar deviasi untuk setiap variabel *input* pada setiap kelas. Berikut persamaan yang digunakan dalam klasifikasi *Gaussian Naïve Bayes* (Brownlee, 2016)

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

$$f(w) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(w - \mu)^2}{\sigma^2}} \quad (2)$$

$$\mu = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (3)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \quad (4)$$

Langkah-langkah proses implementasi metode *Gaussian Naïve Bayes*:

1. Tentukan data latih
2. Menghitung jumlah dan probabilitas, jika terdapat data numerik, maka:
 - a. Tentukan nilai *mean* dan standar deviasi dari masing-masing parameter yang termasuk data numerik.

- b. Tentukan nilai probabilitistik dengan cara menghitung jumlah data yang sesuai dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.

E. *Confusion Matrix*

Confusion Matrix adalah metrik yang digunakan untuk mengevaluasi tingkat akurasi hasil klasifikasi dalam sebuah penelitian. Dalam *Confusion Matrix*, TP (*True Positive*) dan TN (*True Negative*) digunakan ketika klasifikasi menghasilkan hasil yang benar, sedangkan FP (*False Positive*) dan FN (*False Negative*) digunakan ketika terjadi kesalahan dalam klasifikasi tersebut (Han dkk., 2012).

Tabel 1. *Confusion Matrix* 2 Kelas

Aktual	Klasifikasi	
	Yes	No
Yes	<i>True Positive</i>	<i>False Negative</i>
No	<i>False Positive</i>	<i>True Negative</i>

Persamaan untuk pengukuran tingkat akurasi, presisi dan *recall* adalah sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

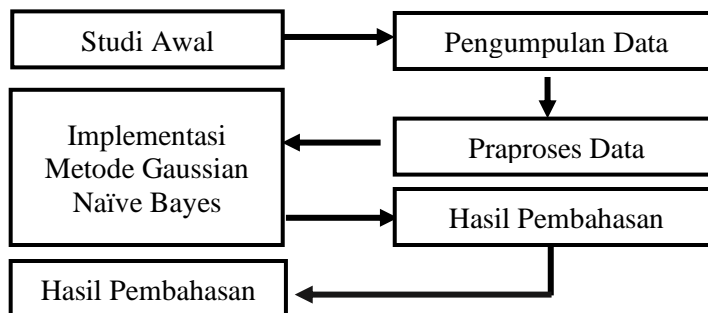
$$Presisi = \frac{TP}{TP + FP} \tag{6}$$

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

F. *K-Fold Cross Validation*

Validasi silang atau *cross validation* dengan *K-Fold* adalah metode yang digunakan untuk mengevaluasi sejauh mana model klasifikasi dapat mengenali data dari kelas yang berbeda. Dalam metode ini, TP (*True Positive*) dan TN (*True Negative*) memberikan informasi tentang kebenaran hasil klasifikasi, sementara FP (*False Positive*) dan FN (*False Negative*) mengindikasikan kesalahan dalam klasifikasi. *Dataset* awal dibagi secara acak menjadi *k subset* yang sama ukurannya, seperti D_1, D_2, \dots, D_k . *Dataset* tersebut kemudian dibagi menjadi data latih dan data uji. Proses pelatihan dan pengujian dilakukan sebanyak *k* kali. Pada setiap iterasi, satu *subset* digunakan sebagai data uji, sedangkan *subset* lainnya digunakan sebagai data pelatihan secara bergantian. Misalnya, pada iterasi pertama, *subset* D_1, D_2, \dots, D_k digunakan sebagai data uji, sementara pada iterasi kedua, *subset* D_2, D_3, \dots, D_k dan D_1 digunakan sebagai data uji, dan seterusnya (Han dkk., 2012).

3. METODE PENELITIAN



Gambar 1. Tahapan Penelitian

Langkah awal dalam penelitian ini adalah mempelajari algoritma *Gaussian Naïve Bayes* dan mengumpulkan informasi tentang penyakit tifus terutama gejala yang terkait dengan penyakit tersebut. Selanjutnya, dilakukan studi literatur, jurnal, dan jurnal yang relevan dengan masalah yang telah dijelaskan, dan menentukan informasi yang dibutuhkan sesuai dengan tujuan penelitian.

Data yang digunakan dalam penelitian ini diambil dari studi yang dilakukan oleh Muhammad Alvinsyah Rizal pada tahun 2020. Total jumlah data yang terkumpul adalah 510 pasien. Data ini terdiri dari 5 atribut, yaitu Demam Tinggi 39-40°C, Sakit Kepala, Badan Lemas, Sakit Perut, dan Penurunan BeratBadan. Atribut-atribut tersebut akan diolah menggunakan algoritma *Gaussian Naïve Bayes* (Rizal, 2022).

Pra-proses data atau *preprocessing data* adalah tahap awal dalam pemrosesan data yang bertujuan untuk memastikan bahwa data penelitian dapat diproses dengan benar oleh algoritma *machine learning*. Dalam penelitian ini, terdapat beberapa metode yang digunakan untuk melakukan pra proses data. Teknik-teknik yang digunakan dalam *preprocessing* di sini adalah transformasi data, standarisasi data dan *imbalanced data*. Implementasi metode *Gaussian Naïve Bayes* memiliki beberapa metode yang membantunya dalam melakukan klasifikasi pada data penelitian ini. Proses dibagi menjadi tiga bagian yaitu pemodelan dan pengujian.

Hasil dari pemodelan dan pengujian perlu dianalisis guna menentukan langkah-langkah yang akan diambil selanjutnya. Analisis tersebut dapat mengungkapkan hasil pengukuran yang kurang baik, sehingga keputusan dapat diambil untuk melakukan perbaikan pada pra proses data atau mengubah metode yang digunakan. Hasil pemodelan dan pengujian juga akan dibandingkan dengan hasil klasifikasi yang diperoleh dari metode *random guess* dan *threshold*. Kedua metode ini berperan sebagai *baseline* atau batas bawah. Perbandingan ini bertujuan untuk menilai apakah metode *Gaussian Naïve Bayes* layak digunakan atau memiliki tingkat akurasi yang memadai.

4. HASIL DAN PEMBAHASAN

A. Pemodelan

Pada penelitian ini, dilakukan pemodelan menggunakan metode *Gaussian Naïve Bayes* dengan tujuan menciptakan model yang baik untuk pengujian data di masa depan. Untuk mencapai evaluasi yang maksimal, penulis menggunakan metode *K-Fold Cross Validation* dalam skenario pemodelan. Skenario *K-Fold Cross Validation* yang digunakan mencakup 2 *fold*, 4 *fold*, 5 *fold*, 8 *fold*, dan 10 *fold*. Setiap *fold* memberikan rata-rata akurasi yang digunakan sebagai acuan dan pembanding terhadap hasil pengujian selanjutnya. Berikut adalah hasil pengukuran yang diperoleh dari pemodelan yang telah dibuat.

Tabel 2. Hasil Pemodelan Menggunakan Skenario *K-Fold Cross Validation*

<i>Fold</i>	Rata-rata Akurasi	Rata-rata Presisi	Rata-rata <i>Recall</i>
k 2	88.87	89.105	88.87
k 4	89.32	89.57	89.32
k 5	89.31	89.63	89.3
k 8	89.61	89.98	89.62
k 10	89.61	90.07	89.62

Berdasarkan pemodelan yang telah dijelaskan, dapat disimpulkan bahwa model *Gaussian Naïve Bayes* memiliki tingkat *akurasi* tertinggi dengan rata-rata sebesar 89,61%, *presisi* tertinggi dengan rata-rata 90,07%, dan *recall* tertinggi dengan rata-rata 89,62%. Ketiga nilai ini ditemukan pada skenario 10 *Fold Cross Validation*.

B. Pengujian

Model yang telah dibangun akan digunakan sebagai referensi dalam pengujian untuk mengevaluasi seberapa baik metode klasifikasi yang digunakan. Sebanyak 20% data akan digunakan sebagai data uji untuk menguji model yang telah dibuat. Perlu dicatat bahwa data uji berbeda dengan data uji validasi. Data uji validasi merupakan data yang digunakan dalam skenario pemodelan dengan metode *K-Fold Cross Validation*, sementara data uji murni adalah data yang tidak boleh disentuh agar menghindari kebocoran data atau data *leakage*. Berikut *confusion matrix* dari klasifikasi pada data uji.

	Negatif	Positif
Actual Negatif	79	4
Actual Positif	2	18

Gambar 4. *Confusion Matrix* Hasil Pengujian

Berdasarkan pengujian yang dilakukan, metode *Gaussian Naïve Bayes* pada data uji mencapai tingkat *akurasi* sebesar 94%, *presisi* rata-rata sebesar 90%, dan *recall* rata-rata sebesar 93%. Hasil pengukuran dari pengujian ini menunjukkan bahwa metode *Gaussian Naïve Bayes* memiliki model yang efektif dalam klasifikasi dan prediksi penyakit demam tifoid pada pasien. Namun, untuk membandingkan keunggulan metode *Gaussian Naïve Bayes* dengan metode lain, diperlukan metode pembandingan. Dalam kasus ini, penulis menggunakan metode *baseline* dan *threshold* sebagai pembandingan terhadap hasil klasifikasi metode *Gaussian Naïve Bayes*.

Baseline dalam pemodelan *machine learning* adalah performa minimum yang dapat dicapai oleh suatu model atau metode dalam memecahkan masalah tertentu. Dalam konteks klasifikasi, *baseline* sering kali merupakan model klasifikasi acak yang mengklasifikasikan data secara acak dengan probabilitas yang sama untuk setiap kelas. *Baseline* digunakan sebagai titik awal untuk membandingkan performa model yang lebih kompleks atau maju. Dalam penelitian ini, penulis menggunakan metode acak untuk membentuk kelas *output* pada data latih dan data uji, dan mengukur tingkat *akurasi* keduanya. Hasilnya, tingkat *akurasi* pada data latih adalah 50% dan pada data uji adalah 38%. Dalam perbandingan dengan pemodelan menggunakan skenario *K-Fold Cross Validation* yang memiliki tingkat *akurasi* 89,61%, metode *Gaussian Naïve Bayes* menunjukkan keunggulan yang signifikan.

Threshold dalam pemodelan *machine learning* adalah nilai batas yang digunakan untuk mengklasifikasikan hasil prediksi model ke dalam kelas positif atau negatif. Dalam penelitian ini, berdasarkan hasil sorting data dengan kondisi tertentu, diasumsikan bahwa pasien dengan gejala "Demam Tinggi 39-40°C" memiliki hasil klasifikasi "Hasil Tifus" yang negatif. Dalam klasifikasi dengan asumsi tersebut, diperoleh nilai *akurasi* sebesar 73% pada pemodelan dan 76% pada pengujian. Namun, jika dibandingkan dengan model yang telah dibangun menggunakan skenario *K-Fold Cross Validation* dengan tingkat *akurasi* 89,61% pada pemodelan dan 94% pada pengujian, dapat disimpulkan bahwa model yang telah dibangun lebih unggul. Hal ini menunjukkan bahwa metode *Gaussian Naïve Bayes* memiliki tingkat *akurasi* yang tinggi dan cocok digunakan untuk klasifikasi dan prediksi pasien penderita demam tifoid.

5. KESIMPULAN

Berdasarkan hasil implementasi metode *Gaussian Naïve Bayes* pada data pasien untuk klasifikasi penyakit demam tifoid yang dilakukan, berikut beberapa poin yang dapat disimpulkan:

- 1) Dari hasil pemodelan menggunakan skenario *K-Fold Cross Validation* diperoleh nilai rata-rata akurasi tertinggi sebesar 89.61%, nilai rata-rata presisi tertinggi sebesar 90.07%, dan nilai rata-rata *recall* tertinggi sebesar 89.62%. Nilai-nilai tersebut diperoleh pada skenario 10 *fold*.
- 2) Dari hasil pengujian model metode *Gaussian Naive Bayes* yang dibangun diperoleh nilai akurasi sebesar 94%, nilai presisi sebesar 90%, dan nilai *recall* sebesar 93%. Hasil ini dapat diketahui bahwa nilai pengujian lebih baik dari nilai pada pemodelan.
- 3) Untuk mengetahui tingkat akurasi model yang dibangun cocok atau tidak, pada penelitian ini dibandingkan dengan metode *Baseline*, dan teknik yang digunakan yaitu *Random Guess* dan *threshold* untuk pembandingan dalam pemodelan dan pengujian data. Dari hasil yang diperoleh nilai pengujian lebih unggul daripada metode *Random Guess* dengan nilai akurasi pemodelan sebesar 49.85% dan pada pengujian 37.86%. Model *Gaussian Naive Bayes* juga unggul terhadap metode *threshold* yang memiliki nilai akurasi pada pemodelan sebesar 73.3% dan pada pengujian sebesar 75.73%.

Saran berdasarkan hasil penelitian untuk peneliti selanjutnya yaitu dapat melakukan pengembangan sebagai berikut:

- 1) Menggunakan metode klasifikasi lain untuk mendapatkan pembandingan dan menemukan metode yang lebih cocok.
- 2) Menerapkan metode *preprocessing* lainnya terutama pada proses untuk mengatasi *imbalanced data*.
- 3) Menggunakan data pasien dari daerah lain untuk mendapatkan variatif data dalam pemodelan.

6. DAFTAR PUSTAKA

- Bhandari, J., & Thada, P. K. (2022). *Typhoid Fever*. <https://www.statpearls.com/point-of-care/30719>.
- Brownlee, J. (2016). *Master Machine Learning Algorithms. Discover How They Work and Implement Them from Scratch*. Jason Brownlee.
- Bustami. (2014). Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah Asuransi. *Jurnal Informatika*, 8(1), 884–898.
- Han, J., Pei, J., & Kamber, M. (2012). *Data Mining: Concepts and Techniques*. Elsevier Inc.
- Hardoon, D. R., Mourão-Miranda, J., Brammer, M., & Shawe-Taylor, J. (2007). Unsupervised analysis of fMRI data using kernel canonical correlation. *NeuroImage*, 37(4), 1250–1259.
- Linson, M., Bresnan, M., Eraklis, A., & Shapiro, F. (1981). Acute gastric volvulus following Harrington rod instrumentation in a patient with Werdnig-Hoffman disease. *Spine*, 6(5), 522–523.
- Rizal M. A. (2022). Analisis Penerapan Algoritma Decision Tree C4.5 Dalam Klasifikasi Pasien Penyakit Tifus (Tifoid) Di Puskesmas Balung Jember. Tugas Akhir. Universitas Muhammadiyah Jember.