

# Optimization of the Naïve Bayes Classifier Algorithm Using Cost-Sensitive Learning to Detect Lung Diseases with an Imbalanced Dataset

Mohammad Zoqi Sarwani<sup>1\*</sup>, Mohamad Khoiron<sup>2</sup>, Muhammad Udin<sup>3</sup>

<sup>1,2,3</sup> Teknik Informatika, Fakultas Teknologi Informasi, Universitas Merdeka Pasuruan, Pasuruan, 67129, Indonesia

## Informasi Artikel

Diterima : 20 Februari 2025  
Revisi : 12 Maret 2025  
Publikasi : 20 Maret 2025

## Kata Kunci:

Naïve Bayes Classifier  
Cost Sensitive Learning  
Penyakit paru-paru  
Klasifikasi  
Optimasi

## ABSTRAK

Penyakit paru-paru merupakan salah satu masalah kesehatan masyarakat global yang terus menjadi perhatian utama di bidang medis. Menurut data dari World Health Organization (WHO), 91% populasi dunia tinggal di daerah dengan kualitas udara yang buruk. Paparan yang terus-menerus terjadi seperti debu, asap rokok, polutan udara, dan zat kimia beracun, dapat menyebabkan resiko terkena penyakit paru-paru. Dalam upaya mengurangi dampak masalah kesehatan pada paru-paru, dan membantu dokter untuk mengklasifikasi penyakit paru-paru. Maka dibutuhkan sebuah metode untuk memprediksi penyakit paru-paru. Naïve Bayes adalah metode teknik klasifikasi yang menggunakan probabilitas dan statistik. Penelitian ini menggunakan dataset sebanyak 30.000 dengan dibagi menjadi data latih dan data uji masing-masing sebesar 80% data latih dan 20% data uji. Hasil dari penelitian ini menunjukkan bahwa optimasi yang dilakukan pada algoritma naïve bayes menggunakan *cost sensitive learning* menghasilkan akurasi sebesar 79,6 % dengan mengalami peningkatan akurasi sebesar 12% dari hasil akurasi sebelumnya tanpa optimasi.

## ABSTRACT

Lung diseases are one of the global public health issues that continue to be a primary concern in the medical field. According to data from the World Health Organization (WHO), 91% of the world's population lives in areas with poor air quality. Continuous exposure to dust, cigarette smoke, air pollutants, and toxic chemicals can increase the risk of developing lung diseases. In efforts to reduce the health impacts on the lungs and assist doctors in classifying lung diseases, a method is needed to predict lung diseases. Naïve Bayes is a classification technique that uses probability and statistics. This research uses a dataset of 30,000, which is divided into training data and testing data, with 80% allocated for training and 20% for testing. The results of this study show that optimization performed on the Naïve Bayes algorithm using cost-sensitive learning achieved an accuracy of 79.6%, which represents a 12% improvement in accuracy compared to the previous result without optimization.

This is an open-access article under the [CC BY-SA](#) license



## \*Penulis Koresponden

Email: zoqi.sarwani@unmerpas.ac.id

Cara sitasi IEEE::

M. Z. Sarwani, M. Khoiron, M. Udin, "Optimization of the Naïve Bayes Classifier Algorithm Using Cost-Sensitive Learning to Detect Lung Diseases with an Imbalanced Dataset" *Journal of Artificial Intelligence and Software Engineering (J-AISE)*, vol. 5, no. 1, pp. 332-338, Maret 2025. doi:10.30811/jaise.v5i1.6474

## 1. PENDAHULUAN

Penyakit paru-paru merupakan masalah kesehatan yang sangat mempengaruhi organ pernapasan utama, yaitu paru-paru. Paru-paru memiliki peran vital dalam tubuh, terutama dalam proses pertukaran oksigen dan karbon dioksida, yang sangat penting untuk kelangsungan hidup. Ketika paru-paru mengalami gangguan, berbagai gejala dapat muncul, yang dapat bervariasi mulai dari sesak napas hingga penurunan kemampuan tubuh dalam menjalankan aktivitas sehari-hari. Jika tidak ditangani dengan baik, penyakit paru-paru dapat menyebabkan komplikasi serius yang mengganggu kualitas hidup dan bahkan mengancam jiwa.

Menurut data dari *World Health Organization (WHO)*, 91% populasi dunia tinggal di daerah dengan kualitas udara yang buruk [1]. Paparan yang terus-menerus terjadi seperti debu, asap rokok, polutan udara, dan zat kimia beracun, dapat menyebabkan resiko terkena penyakit paru-paru. Penyakit Paru Obstruktif Kronis (PPOK) menjadi salah satu penyakit paru-paru yang paling umum di Indonesia [2]. Data menunjukkan bahwa ada sekitar 10,9 juta orang di Indonesia menderita PPOK pada tahun 2017. Penyakit paru-paru terus menjadi perhatian utama di bidang medis. Meskipun telah ada kemajuan dalam diagnosis dan pengobatan, masih ada permasalahan yang belum terselesaikan dalam penanganan penyakit paru-paru.

Meskipun perkembangan dalam bidang diagnosis dan pengobatan penyakit paru-paru telah mencapai kemajuan, masalah penanganannya masih menjadi tantangan besar. Keterlambatan dalam diagnosis atau kurangnya deteksi dini sering kali menyebabkan pasien mengalami kondisi yang lebih parah. Oleh karena itu, inovasi dalam sistem deteksi otomatis dan prediksi penyakit paru-paru menjadi sangat penting.

Dalam upaya mengatasi masalah ini, penggunaan algoritma klasifikasi dalam bidang kesehatan, seperti *Naïve Bayes*, telah menunjukkan potensi yang menjanjikan. Algoritma *Naïve Bayes* adalah metode teknik klasifikasi yang menggunakan probabilitas dan statistik [3]. Metode *Naïve Bayes* telah banyak digunakan oleh para peneliti sebelumnya dengan berbagai dataset. Pada penelitian yang dilakukan oleh [4] menerapkan algoritma *Naïve Bayes* untuk memprediksi status pasien Covid-19 di Indonesia dengan nilai presisi 92%, recall 88.72% dan akurasi 96.67%. Studi lain oleh [5] menerapkan metode *Naïve Bayes* dalam memprediksi penyakit jantung pada dataset 300 baris data dengan nilai akurasi 81%. Studi lain oleh [6] menerapkan metode *Naïve Bayes* untuk mendiagnosa penyakit ISPA dengan nilai tingkat akurasi sebesar 92.3%. Studi lain oleh [7] menerapkan algoritma *Naïve Bayes* untuk mengklasifikasi kanker payudara dengan mencapai nilai akurasi 96.9%.

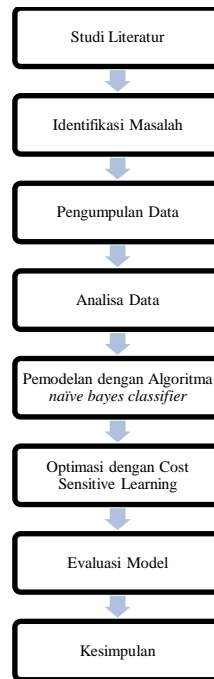
Namun, dalam banyak kasus, data yang tidak seimbang menjadi masalah utama yang dapat mempengaruhi kinerja model klasifikasi. Ketidakseimbangan data ini terjadi ketika jumlah sampel dalam suatu kelas jauh lebih banyak atau lebih sedikit dibandingkan dengan kelas lainnya. Kondisi ini sering menyebabkan model lebih cenderung untuk memprediksi kelas mayoritas, sehingga mengurangi akurasi dan keandalan dalam memprediksi kelas minoritas [8]. Sehingga terdapat banyak penelitian yang mencoba menyelesaikannya dengan banyak pendekatan salah satunya adalah dengan *cost sensitive learning* [9], [10]. Metode *cost sensitive learning* dapat menyelesaikan berbagai permasalahan baik dari data tidak seimbang hingga melakukan optimasi seperti yang dilakukan oleh [11]. Dimana pada penelitian [9] menggunakan *cost sensitive learning* untuk menyelesaikan permasalahan klasifikasi. Peneliti lain menggunakan *cost sensitive learning* untuk melakukan prediksi dibidang keuangan [11].

Berdasarkan hal tersebut yang telah dijelaskan, dapat dilihat bahwa algoritma *Naïve Bayes Classifier* memiliki kelemahan dalam menangani data tidak seimbang dilihat dari perbedaan antara nilai *precision* dan *recall*. Sehingga pada penelitian ini akan melakukan optimasi pada algoritma *naïve bayes classifier* dengan menggunakan *cost sensitive learning*. Selain itu, *cost sensitive learning* yang diterapkan pada penelitian ini diharapkan dapat menangani permasalahan data tidak seimbang pada kasus deteksi penyakit paru-paru.

## 2. METODE

### 2.1 Alur Penelitian

Pada penelitian ini terdapat beberapa tahapan yang akan dilakukan. Tahapan tersebut dimulai dengan melakukan studi literatur, identifikasi masalah, pengumpulan data, analisa data, pemodelan dan evaluasi model. Pada gambar 1 ditunjukkan alur dari penelitian ini.



Gambar 1. Alur penelitian

Studi literatur merupakan tahapan yang dilakukan oleh peneliti untuk mengumpulkan referensi yang digunakan dalam memecahkan permasalahan yang diambil oleh peneliti. Proses identifikasi masalah merupakan tahapan untuk mendapatkan gambaran masalah yang akan dipecahkan dalam penelitian ini. Proses pengumpulan data merupakan proses atau tahapan yang digunakan untuk mengumpulkan data mentah dan *feature engineering* yakni proses untuk melakukan perbaikan data serta memastikan data tidak memiliki nilai yang kosong untuk setiap fiturnya. Selanjutnya akan di proses pada tahap analisa data sehingga dapat ditentukan parameter yang tepat dengan melakukan proses analisa statistik deskriptif untuk mendapatkan gambaran hubungan antar parameter yang digunakan pada penelitian ini. Data yang telah melewati proses analisa data selanjutnya akan di proses oleh algoritma *Naïve Bayes Classifier* yang dioptimasi menggunakan *cost sensitive learning*. Proses terakhir adalah melakukan evaluasi model sehingga dapat mengetahui seberapa baik model yang dihasilkan oleh algoritma *naïve bayes classifier* yang telah dioptimasi dalam menyelesaikan permasalahan pada penelitian ini.

## 2.2 Studi Literatur

Pada tahap ini peneliti melakukan pengumpulan referensi yang digunakan untuk memecahkan permasalahan yang diambil oleh peneliti. Referensi yang digunakan pada penelitian ini diambil dari jurnal, buku, dan sumber lain yang terpercaya. Referensi yang diambil oleh peneliti digunakan untuk menguatkan teori yang dituangkan oleh peneliti dalam latar belakang sehingga peneliti dapat mengajukan solusi yang berjudul optimasi algoritma *naïve bayes classifier* menggunakan *cost sensitive learning* untuk klasifikasi penyakit paru-paru.

## 2.3. Identifikasi Masalah

Pada penelitian ini proses identifikasi masalah digunakan untuk mendapat gambaran yang detail tentang masalah yang dipecahkan dalam penelitian ini. Penyakit paru-paru merupakan salah satu penyakit yang menyebabkan kematian. Masyarakat sering menganggap penyakit paru-paru seperti biasa karena gejala yang dialami mirip dengan penyakit biasa. Ketidaktahuan tersebut yang menjadi salah satu penyebab penyakit paru-paru mengakibatkan kematian. Dibutuhkan sebuah platform atau teknologi yang dapat diakses dengan mudah oleh seluruh masyarakat serta dapat memberikan identifikasi awal kepada masyarakat tentang gejala yang dialaminya merupakan penyakit paru-paru atau bukan sehingga dapat mencegah sejak dini.

## 2.4. Pengumpulan Data

Pada penelitian ini menggunakan data publik yang diambil dari situs kaggle berupa file excel. Data tersebut berjumlah 30.000 baris data dengan 10 kolom atribut yaitu Usia, Jenis Kelamin, Merokok, Bekerja, Rumah Tangga, Aktivitas Begadang, Aktivitas Olahraga, Asuransi, Penyakit Bawaan, dan Hasil. Parameter

Hasil merupakan parameter yang akan dijadikan sebagai parameter target atau variable dependen. Pada tabel 1 ditunjukkan contoh dari data yang akan diproses.

Tabel 1. Dataset penyakit paru-paru

No.	Umur	Jenis Kelamin	Perokok	Bekerja	Rumah Tangga	Aktivitas Begadang	Aktivitas Olahraga	Asuransi	Penyakit Bawaan	Hasil
1.	Muda	Pria	Aktif	Tidak	Ya	Ya	Sering	Tidak	Tidak	Tidak
2.	Muda	Wanita	Bwaan	Ya	Tidak	Tidak	Sering	Tidak	Ya	Ya
...										
29.000	Tua	Wanita	Pasif	Ya	Tidak	Tidak	Sering	Tidak	Ya	Tidak
30.000	Muda	Wanita	Pasif	Tidak	Ya	Tidak	Sering	Tidak	Tidak	Ya

2.5 Analisa Data

Analisa data merupakan tahapan yang digunakan oleh peneliti untuk menentukan parameter yang dijadikan sebagai input dari algoritma *naïve bayes classifier* serta untuk melakukan perbaikan fitur atau *feature engineering*. Pada penelitian ini proses analisa data dilakukan dengan beberapa tahapan yang dimulai dari menganalisa korelasi antar parameter, mendeteksi data seimbang atau tidak berdasarkan parameter target, menangani *missing value*, dan mengubah nilai dari parameter agar setiap nilai dari parameter dapat diproses oleh algoritma *naïve bayes classifier*.

2.6 Pemodelan dengan Algoritma *Naïve Bayes Classifier*

Pemodelan merupakan tahapan yang digunakan untuk mendapatkan pola terbaik berdasarkan data. Pada penelitian ini algoritma yang digunakan untuk memodelkan data adalah algoritma *Naïve Bayes Classifier*. Algoritma *Naïve Bayes Classifier* bekerja berdasarkan teorema bayes yang menggunakan probabilitas dalam menentukan hasil [12]. Terdapat berbagai macam algoritma *Naïve Bayes Classifier* antara lain *Gaussian Naïve Bayess*, *Multinomial Naïve Bayess*, *Bernauli Naïve Bayess*. Pada penelitian ini jenis algoritma *Naïve Bayes Classifier* yang digunakan adalah *Bernauli Naïve Bayess*. Algoritma ini sangat sesuai digunakan ketika fitur data bersifat biner, misalnya, nilai 0 atau 1, atau benar/salah. Selain itu, algoritma ini juga efektif untuk melakukan klasifikasi data yang hanya memiliki 2 klas. Sehingga pemilihan algoritma *Bernouli Naïve Bayess* sangat cocok karena data yang digunakan pada penelitian ini memiliki fitur dan target klas bersifat biner. Cara Kerja *Bernoulli Naïve Bayes* [13] adalah sebagai berikut :

a. Probabilitas A Priori

Hitung probabilitas a priori dari setiap kelas ( $C_k$ ) berdasarkan proporsi kemunculannya dalam dataset.

$$P(C_k) = \frac{\text{jumlah sampel dengan kelas } C_k}{\text{total jumlah sampel}} \tag{1}$$

b. Probabilitas Likelihood

Untuk setiap fitur  $x_i$  dan setiap kelas  $C_k$ , hitung probabilitas ( $P(x_i = 1|C_k)$ ) dan ( $P(x_i = 0|C_k)$ )

$$P(x_i = 1|C_k) = \frac{\text{jumlah sampel dengan } x_i=1 \text{ dan kelas } C_k}{\text{jumlah sampel dengan kelas } C_k} \tag{2}$$

$$P(x_i = 0|C_k) = 1 - P(x_i = 1|C_k) \tag{3}$$

c. Prediksi

Diberikan data baru  $X = (x_1, x_2, \dots, x_n)$ , hitung probabilitas posterior untuk setiap kelas  $C_k$  menggunakan rumus:

$$P(C_k|X) \propto P(C_k) \times \prod(P(x_i|C_k)) \tag{4}$$

Di mana:

$$P(x_i|C_k) = P(x_i = 1|C_k) \text{ jika } x_i = 1$$

$$P(x_i|C_k) = P(x_i = 0|C_k) \text{ jika } x_i = 0$$

d. Klasifikasi

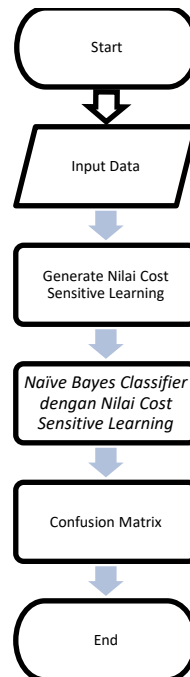
Kelas dengan probabilitas posterior tertinggi dipilih sebagai prediksi.

Contoh Penggunaan

Misalkan kita memiliki dataset dengan fitur biner yang mewakili adanya atau tidaknya suatu atribut, seperti gejala penyakit, dan kita ingin memprediksi apakah seseorang menderita penyakit tertentu (target biner: sakit atau tidak sakit). Selain itu *Bernoulli Naïve Bayes* memiliki keunggulan antara lain : efisien untuk data biner, sederhana dan mudah diimplementasikan dan kinerja baik dalam kasus dengan fitur yang bernilai biner.

2.7 Optimasi dengan *Cost Sensitive Learning*

Optimasi merupakan cara yang digunakan untuk meningkatkan akurasi dari sebuah model. Pada penelitian ini teknik atau metode optimasi yang digunakan adalah *cost sensitive learning*. Metode ini bekerja dengan cara memberikan bobot pada klas minoritas sehingga mampu memberikan akurasi yang optimal dan dapat menangani masalah data tidak seimbangan. Penentuan bobot untuk setiap klas bisa dilakukan dengan beberapa cara yakni secara manual atau menggunakan library yang disediakan oleh *scikit-learn*. Pada penelitian ini penentuan bobot dilakukan dengan memanfaatkan fitur yang terdapat pada *library scikit-learn* yang menggunakan modul *computed\_sample\_weight*. Pada gambar 2 menunjukkan proses algoritma *Naïve Bayes Classifier* dioptimasi dengan *cost sensitive learning*.



Gambar 2. Flowchat Metode Optimasi Naive bayes dengan cost sensitive learning

## 2.8 Evaluasi Model

Evaluasi model merupakan tahapan yang digunakan untuk mengukur seberapa baik metode atau algoritma klasifikasi dalam mendeteksi penyakit paru-paru. Pada penelitian ini metode evaluasi model yang digunakan adalah metode *confusion matrix*. *Confusion matrix* memberikan gambaran tentang bagaimana performa model dalam hal prediksi benar dan salah, serta kesalahan tipe apa yang dibuat oleh model [14]. *Confusion matrix* biasanya terdiri dari empat komponen utama: True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).

### Komponen *Confusion Matrix*

#### True Positive (TP)

True Positive (TP) adalah jumlah sampel yang benar-benar positif dan diprediksi positif oleh model. Ini menunjukkan bahwa model berhasil mendeteksi kelas positif dengan benar.

#### True Negative (TN)

True Negative (TN) adalah jumlah sampel yang benar-benar negatif dan diprediksi negatif oleh model. Ini menunjukkan bahwa model berhasil mendeteksi kelas negatif dengan benar.

#### False Positive (FP)

False Positive (FP) adalah jumlah sampel yang sebenarnya negatif tetapi diprediksi positif oleh model. Ini dikenal juga sebagai 'Type I Error'.

#### False Negative (FN)

False Negative (FN) adalah jumlah sampel yang sebenarnya positif tetapi diprediksi negatif oleh model. Ini dikenal juga sebagai 'Type II Error'.

Berikut formula evaluasi dengan *Confusion Matrix*

#### 1. Accuracy

Accuracy adalah proporsi prediksi yang benar (baik positif maupun negatif) dari keseluruhan prediksi.

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (5)$$

## 2. Precision

*Precision* adalah proporsi prediksi positif yang benar dari semua prediksi positif.

$$Precision = \frac{TP}{(TP+FP)} \quad (6)$$

## 3. Recall (Sensitivity)

*Recall*, juga dikenal sebagai *Sensitivity* atau *True Positive Rate (TPR)*, adalah proporsi prediksi positif yang benar dari semua kasus positif aktual.

$$Recall = \frac{TP}{(TP + FN)}$$

## 4. F1-Score

*F1-Score* adalah harmonisasi rata-rata dari *precision* dan *recall*, memberikan keseimbangan antara keduanya.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

## 3. HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas hasil yang telah dilakukan berdasarkan rancangan yang dibuat pada bab sebelumnya. Hasil yang dibahas pada bab ini meliputi analisa data, *naïve bayess classifier*, optimasi *naïve bayess classifier* dengan *cost sensitive learning* dan evaluasi model menggunakan *confusion matrix*.

### 3.1 Analisa Data

Pada penelitian ini, data yang digunakan adalah *lung disease* dataset yang diambil dari Kaggle.com. Data ini berupa data tabular yang berisi berbagai fitur yang terkait dengan penyakit paru-paru. Sebelum melakukan analisis, dilakukan tahap pra-pemrosesan data seperti pengisian nilai yang hilang, label kategorikal, dan normalisasi data.

Setelah pra-pemrosesan, dilakukan eksplorasi data untuk memahami distribusi dan hubungan antar variabel. Dari eksplorasi yang dilakukan didapat dua analisis yaitu analisis deskriptif dan analisis korelasi. Dari hasil analisis deskriptif mayoritas pasien dalam dataset ini berusia antara 40-60 tahun sedangkan analisis korelasi, fitur-fitur seperti `umur`, `aktivitas perokok`, dan `penyakit bawaan` memiliki korelasi yang signifikan dengan variabel target `hasil`.

### 3.2 Naïve Bayes Classifier

Algoritma *Naïve Bayes* adalah salah satu algoritma yang digunakan untuk klasifikasi dalam penelitian ini. Algoritma ini bekerja berdasarkan Teorema Bayes dengan asumsi bahwa fitur-fitur yang digunakan adalah independen satu sama lain.

Pada penelitian ini sebelum diproses oleh algoritma *Naïve Bayes Classifier*, data akan dibagi menjadi data latih dan data uji dengan proporsi pembagian data sebesar 80% untuk data latih dan 20% untuk data uji. Model *Naïve Bayes Classifier* diterapkan pada dataset dan dilakukan evaluasi menggunakan metrik-metrik seperti akurasi, presisi, recall, dan F1-score. Berikut adalah hasil dari model *Naïve Bayes Classifier* didapatkan akurasi sebesar 67%, presisi sebesar 82,2%, recall sebesar 70,3% dan F1-Score sebesar 75,7%.

### 3.3 Optimasi Naïve Bayes Classifier dengan Cost Sensitive Learning

Untuk meningkatkan performa dari model *Naïve Bayes Classifier*, dilakukan optimasi dengan menggunakan teknik *Cost Sensitive Learning*. Teknik ini memperhitungkan biaya dari kesalahan klasifikasi yang berbeda, yang dalam konteks ini, *misclassification cost* untuk *False Negative (FN)* lebih tinggi dibandingkan dengan *False Positive (FP)* karena risiko kesehatan yang lebih besar. Setelah menerapkan *Cost Sensitive Learning* pada model *Naïve Bayes Classifier*, diperoleh hasil akurasi sebesar 79,6%, presisi sebesar 86,2%, *recall* sebesar 82,5%, dan *F1-Score* sebesar 84,3%. Peningkatan performa ini menunjukkan bahwa model yang dioptimalkan mampu mengurangi jumlah *False Negative* dengan tetap menjaga tingkat *False Positive* yang rendah.

### 3.4 Confusion Matrix

*Confusion Matrix* digunakan untuk memberikan gambaran yang lebih mendetail mengenai performa model. Berikut adalah *Confusion Matrix* sebelum dan sesudah optimasi:

Tabel 2. Analisa nilai sebelum optimasi

Kategori	True Positive	True Negative	False Positive	False Negative
Total	3.137	863	678	1322

Table 3. Analisa nilai setelah optimasi

Kategori	True Positive	True Negative	False Positive	False Negative
Total	3.292	1489	523	696

Dari *Confusion Matrix* di atas, dapat dilihat bahwa setelah optimasi, jumlah *False Negative* (696) lebih sedikit dibandingkan sebelum optimasi (1322), sementara jumlah *False Positive* menurun sedikit dari 678 menjadi 523. Hal ini menunjukkan bahwa model yang dioptimalkan lebih efektif dalam mendeteksi kasus positif, yang sangat penting dalam konteks deteksi penyakit paru-paru.

#### 4. KESIMPULAN

Pada penelitian ini yang melakukan optimasi algoritma *naïve bayes classifier* menggunakan *cost sensitive learning* menghasilkan akurasi sebesar 79,6% dengan mengalami peningkatan akurasi sebesar 12% dari hasil akurasi sebelumnya tanpa optimasi. Selain itu, pada hasil *confusion matrix* yang ditunjukkan pada bab sebelumnya menunjukkan bahwa optimasi dengan *cost sensitive learning* dapat menangani data yang tidak seimbang. Dimana data yang digunakan sebesar 30.000 yang dibagi mejadi data latih dan data uji masing-masing sebesar 80% untuk data latih dan 20% untuk data uji.

Berdasarkan hal tersebut menunjukkan bahwa penelitian ini dapat memberikan hasil yang optimal dalam mendeteksi penyakit paru-paru dengan data yang tidak seimbang dengan melakukan teknik optimasi. Sehingga pada penelitian berikutnya dapat menguji pendekatan solusi yang diberikan pada penelitian ini dengan menerapkan pada kasus yang berbeda. Selain itu, penelitian ini masih berpotensi untuk dilakukan improvisasi guna untuk meningkatkan akurasi.

#### REFERENSI

- [1] V. A. Ulfa, A. Asnifatima, and A. Fathimah, "Faktor-Faktor yang Berhubungan dengan Kejadian Sick Building Syndrome (SBS) pada Karyawan RSIA Pasutri Bogor Tahun 2020," vol. 5, no. 5, pp. 428–434, 2022, doi: 10.32832/pro.
- [2] M. Y. Haffandi, E. Haerani, F. Syafria, and L. Oktavia, "Klasifikasi Penyakit Paru-Paru Dengan Menggunakan Metode Naïve Bayes Classifier," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 5, no. 2, p. 176, Dec. 2022, doi: 10.37600/tekinkom.v5i2.649.
- [3] M. Sabransyah, Y. Novia Nasution, F. Deny, and T. Amijaya, "Aplikasi Metode Naïve Bayes dalam Prediksi Risiko Penyakit Jantung Naïve Bayes Method for a Heart Risk Disease Prediction Application," *Jurnal EKSPONENSIAL*, vol. 8, no. 2, 2017.
- [4] D. Yanti Liliana, H. Maulana, and A. Setiawan, "Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naïve Bayes," 2021. [Online]. Available: [www.kaggle.com](http://www.kaggle.com)
- [5] A. Atthohiroh, R. Ayu, And S. Maharani, "Penerapan Metode Naïve Bayes Untuk Memprediksi Penyakit Jantung," *JURNAL TEKNISI*, vol. 3, no. 1, p. 8, Feb. 2023, doi: 10.54314/teknisi.v3i1.1252.
- [6] F. Ramadhana, F. Fauziah, and W. Winarsih, "Aplikasi Sistem Pakar untuk Mendiagnosa Penyakit ISPA menggunakan Metode Naïve Bayes Berbasis Website," *STRING (Satuan Tulisan Riset dan Inovasi Teknologi)*, vol. 4, no. 3, p. 320, Apr. 2020, doi: 10.30998/string.v4i3.5441.
- [7] H. Oktavianto and R. P. Handri, "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naïve Bayes," *INFORMAL: Informatics Journal*, vol. 4, no. 3, p. 117, Jan. 2020, doi: 10.19184/isj.v4i3.14170.
- [8] Z. Wu, H. Zhang, J. Guo, Y. Ji, and M. Pecht, "Imbalanced bearing fault diagnosis under variant working conditions using cost-sensitive deep domain adaptation network," *Expert Syst Appl*, vol. 193, p. 116459, May 2022, doi: 10.1016/j.eswa.2021.116459.
- [9] O. Volk and G. Singer, "Adaptive Cost-Sensitive Learning in Neural Networks for Misclassification Cost Problems," Nov. 2021, [Online]. Available: <http://arxiv.org/abs/2111.07382>
- [10] S. Mählkvist, J. Ejenstam, and K. Kyprianidis, "Cost-Sensitive Decision Support for Industrial Batch Processes," *Sensors*, vol. 23, no. 23, p. 9464, Nov. 2023, doi: 10.3390/s23239464.
- [11] S. A.-D. Safi, P. A. Castillo, and H. Faris, "Cost-Sensitive Metaheuristic Optimization-Based Neural Network with Ensemble Learning for Financial Distress Prediction," *Applied Sciences*, vol. 12, no. 14, p. 6918, Jul. 2022, doi: 10.3390/app12146918.
- [12] H. Chen, S. Hu, R. Hua, and X. Zhao, "Improved Naïve Bayes classification algorithm for traffic risk management," *EURASIP J Adv Signal Process*, vol. 2021, no. 1, pp. 1–12, Dec. 2021, doi: 10.1186/S13634-021-00742-6/TABLES/5.
- [13] M. Artur, "Review the performance of the Bernoulli Naïve Bayes Classifier in Intrusion Detection Systems using Recursive Feature Elimination with Cross-validated selection of the best number of features," *Procedia Comput Sci*, vol. 190, pp. 564–570, Jan. 2021, doi: 10.1016/J.PROCS.2021.06.066.
- [14] H. M and S. M.N, "A Review on Evaluation Metrics for Data Classification Evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/IJDKP.2015.5201.