

KLASIFIKASI PENYAKIT PARU-PARU DENGAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER

Muhammad Yusril Haffandi¹⁾, Elin Haerani²⁾, Fadhilah Syafria³⁾, Lola Oktavia⁴⁾

^{1,2,3,4} Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim

email: 1750115256@students.uin-suska.ac.id, elin.haerani@uin-suska.ac.id,

fadhilah.syafria@uin-suska.ac.id, lola.oktavia@uin-suska.ac.id

Abstract



The lungs are one of the organs of the human body that are very important in the process of respiration. There are several types of lung diseases, including Asthma, Bronchitis, Dyspnea, Pneumonia, COPD, and Tuberculosis. There are difficulties in the classification process, because the symptoms shown by sufferers have similarities between diseases. The purpose of this research is to classify lung disease using the Naive Bayes Classifier method. The choice of this method is because it only requires a small amount of training data to determine the estimated parameters needed in the classification process. This research was conducted at the Regional General Hospital Major General HA Thalib City of Sungai Full from August 3 2022 to September 3 2022. The data taken was in the form of medical records of lung disease patients from July to August as many as 134 patient data containing 19 symptoms disease and 6 disease diagnoses. From the test results using the Rapidminer application and data separation in the form of 34 testing data and 100 training data with a data comparison of 7:3, an accuracy value of 97.06 was obtained.

Keywords: Lung Disease, Naïve Bayes Classifier Method, Symptoms, Classification, Data Mining

1. PENDAHULUAN

Paru-paru adalah sebuah organ yang memiliki fungsi respirasi. Respirasi adalah proses pelepasan energi yang tersimpan dalam zat sumber energi dengan proses kimiawi yang menggunakan O₂, proses pengambilan O₂ untuk memecah senyawa organik menjadi CO₂, H₂O dan energi [1]. Penyakit paru-paru merupakan penyakit yang sering terjadi pada manusia. Penyebab terjadinya penyakit paru-paru biasanya dikarenakan menghirup udara yang telah tercemar, baik dari debu, asap, virus, maupun bakteri yang mengakibatkan infeksi pada saluran pernafasan [2]. Mulai dari bayi hingga orang dewasa bisa terjangkit penyakit paru-paru. Tentunya penyakit ini tidak mudah untuk disembuhkan.

Demi mempertimbangkan pentingnya mengetahui jenis penyakit paru-paru menurut jumlah kasus paru-paru di Indonesia dan gejala yang dialami pasien, salah satu penyakit paru-paru yang mematikan adalah PPOK. Menurut World Health Organization yang dirilis pada

tahun 2008 oleh WHO, PPOK diperkirakan menjadi penyebab kematian ketiga pada tahun 2030, dan jumlah kecacatan akibat PPOK diproyeksikan meningkat dari urutan ke-9 menjadi urutan ke-5 pada tahun 2020 [3].

Data mining adalah proses pencarian aturan atau pola dari sejumlah data yang memiliki kemungkinan sangat tinggi dalam menemukan informasi baru [4]. Data mining sering disebut sebagai Knowledge Discovery Databases (KDD) karena terutama digunakan untuk menemukan informasi dari basis data yang berskala besar [5]. Salah satu teknik yang penting dalam data mining yaitu klasifikasi. Klasifikasi adalah proses untuk mencari suatu pola yang menggambarkan atau membedakan tiap kelas data dengan tujuan untuk menyimpulkan kelas objek yang peruntukannya tidak diketahui [6].

Umumnya masyarakat awam hanya mengetahui penyakit paru-paru sebagai asma, namun terdapat berbagai jenis penyakit paru-paru seperti bronkhitis, asma, kanker paru-paru,

pneumonia, penyakit paru obstruktif kronik (PPOK), tuberkulosis (TB), efusi pleura, emfisema, dan pneumotoraks [7]. Penyakit paru yang disebutkan tentunya memiliki gejala dan penanganan yang berbeda tergantung jenis penyakit yang diderita. Oleh karena itu, penderita penyakit harus benar-benar mengetahui penyakit yang dideritanya. Gangguan ini dapat berupa sesak napas yang menyebabkan kesulitan melakukan aktivitas dan menyebabkan hipoksia pada individu yang terkena. Gangguan ini bahkan dapat menyebabkan kematian [8].

Seiring perkembangan zaman, transportasi menjadi sarana dan alat kendaraan dalam melakukan perjalanan. Meningkatnya jumlah volume kendaraan juga meningkatkan masalah pencemaran udara. Dengan kemajuan teknologi juga dalam memberikan inovasi, tentunya semakin banyak pabrik-pabrik industri yang beroperasi. Limbah dari pabrik industri, polusi dari alat transportasi, asap akibat pembakaran yang terjadi setiap harinya tentu akan berimbas pada pencemaran udara. Dari banyaknya asap yang dihasilkan dalam industri pabrik juga membuat lingkungan sekitar menjadi tercemar dan dapat mengakibatkan pemanasan global [9]. Dalam satu sisi memberikan dampak positif terhadap perkembangan ekonomi, namun dampak negatifnya berupa pencemaran udara dari peningkatan emisi kendaraan [10].

Dalam proses data mining, terdapat berbagai metode yang diterapkan pada proses klasifikasi berdasarkan faktor ketidakpastian (*uncertainty*), dan salah satunya adalah metode Naive Bayes. Teorema Naive Bayes banyak digunakan dalam studi kasus yang melibatkan diagnosis statistik dan berhubungan dengan probabilitas serta kemungkinan dari suatu penyakit dengan gejala yang terkait [11]. Metode ini juga memiliki beberapa kelebihan atau keuntungan yaitu metode ini hanya membutuhkan *Data Training* yang sedikit untuk menentukan kemungkinan suatu parameter yang dibutuhkan dalam proses klasifikasi [12]. Selain itu, metode Naive Bayes mudah untuk diimplementasikan dan metode ini juga

memberikan hasil yang baik bahkan dalam berbagai kasus di dunia nyata yang bersifat kompleks. Penelitian yang dilakukan oleh Indah [13] juga menyebutkan bahwa metode Naive Bayes memiliki nilai akurasi tinggi jika diimplementasikan dengan menggunakan data yang berskala besar.

Berdasarkan penelitian sebelumnya yang dilakukan oleh Benny dan Juniati [14] mengenai klasifikasi penyakit paru berdasarkan citra x-ray thorax dengan menggunakan metode fractal box counting, didapatkan nilai akurasi sebesar 79,65%. Yang menjadikan perbedaan terhadap penelitian terdahulu terletak pada sumber data, yang mana sumber pengetahuan yang diperoleh berasal dari gejala penyakit yang diderita oleh pasien.

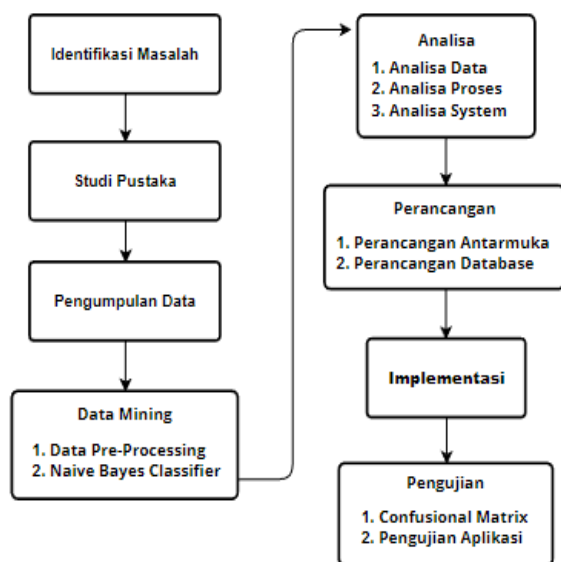
Penelitian ini memberikan kontribusi dalam mengklasifikasikan penyakit paru yang diderita pasien dan mempermudah dalam melakukan klasifikasi penyakit paru-paru menggunakan metode Naive Bayes Classifier.

2. METODE PENELITIAN

Metode penelitian yang digunakan yaitu metode Naive Bayes Classifier. Metode Naive Bayes Classifier adalah pengklasifikasi probabilitas yang sederhana berdasarkan Teorema Bayes. Naive Bayes Classifier merupakan algoritma pembelajaran yang terawasi (*supervised*), dan metode ini mengasumsikan bahwa efek fitur tertentu tidak bergantung dengan fitur lainnya yang karena inilah metode ini disebut juga dengan metode naïf (*naïve*). Adapun langkah-langkah penelitian terlihat pada gambar 1.

Gambar 1 merupakan tahapan dalam melakukan penelitian yang berurut yaitu:

1. Identifikasi Masalah
Tujuan dari fase ini adalah untuk mendapatkan gambaran tentang permasalahan yang ada.
2. Studi Pustaka
Pada tahap ini data dan informasi terkait penelitian dikumpulkan melalui berbagai media, seperti jurnal, buku, internet atau dari penelitian terkait.



Gambar 1. Alur Penelitian

3. Pengumpulan Data

Proses pengumpulan data dilakukan saat melakukan penelitian di RSUD Mayjen H.A. Thalib dari tanggal 3 Agustus sampai 3 September 2022. Data yang diperoleh berupa rekam medis pasien yang mengidap penyakit paru-paru.

4. Data Mining

Pada tahap ini, dilakukan beberapa tahapan Pre-Processing. Pre-Processing adalah proses mengubah atau mengolah suatu data yang tidak tertata menjadi data yang lebih tertata [15]. Tujuan dilakukan Pre-Processing yaitu untuk mengecilkan kegagalan, data yang tidak konsisten, atau data duplikat [16].

5. Analisa

Tujuan dilakukan analisa yaitu untuk menyelaraskan data, tampilan dan proses dari penelitian.

6. Perancangan

Tahap ini dilakukan perancangan terhadap antarmuka dan database. Perancangan antarmuka sebagai bentuk gambaran aplikasi yang dibangun dan perancangan database sebagai bentuk data dalam aplikasi yang dibangun.

7. Implementasi

Untuk menjalankan sistem yang sedang dirancang, dibutuhkan perangkat yang memadai dalam menjalankan dan pengkodean sistem. Maksud dilakukannya pengkodean untuk merubah rancangan yang telah dibuat agar dapat dikenali oleh komputer [17].

8. Pengujian

Untuk menguji keakuratan sistem dalam melakukan perhitungan, dilakukan pengujian untuk menghitung akurasi dan performa dari sistem. Salah satu tools yang digunakan yaitu Rapidminer. Rapid miner adalah tools pengolahan data yang bersifat open core dan biasanya digunakan dalam pengolahan data pada machine learning, text mining, data mining, dan predictive analytic [18].

3. HASIL DAN PEMBAHASAN

3.1. Persiapan Data

Data rekam medis pasien yang digunakan pada penelitian ini diambil di Rumah Sakit Umum Daerah Mayjen H.A. Thalib Kota Sungai Penuh. Total data yang diperoleh secara keseluruhan yang diambil dari bulan Juli sampai Agustus berjumlah 134 data dengan 24 atribut dan 1 kelas.

3.2. Data Pre-Processing

Sebelum memasuki proses mining terhadap data, dilakukan pembersihan (cleaning) dan transformasi (transformation) pada data yang akan diolah.

Bentuk perubahan dari Data Transformation akan ditampilkan pada tabel 1.

Tabel 1. Data Penyakit

Atribut	Binominal
Ya	1
Tidak	0

Setelah melalui proses Data Transformation, bentuk dari atribut gejala penyakit akan ditampilkan pada tabel 2.

Tabel 2. Data Gejala Penyakit

No	Kode Gejala	Gejala
1.	G1	Sesak Nafas
2.	G2	Batuk
3.	G3	Batuk Berdarah
4.	G4	Berdahak
5.	G5	Demam
6.	G6	Lemas
7.	G7	Nafsu Makan Berkurang
8.	G8	Mual
9.	G9	Muntah
10.	G10	BAB
11.	G11	BAK
12.	G12	Riwayat Penyakit Asma
13.	G13	Riwayat Penyakit Stroke
14.	G14	Riwayat Penyakit Tuberkulosis
15.	G15	Sakit Kepala
16.	G16	Nyeri Ulu Hati
17.	G17	Nyeri Dada
18.	G18	Berat Badan Menurun
19.	G19	Keringat Malam

Bentuk dari nama penyakit akan ditampilkan pada tabel 3.

Tabel 3. Data Penyakit

No	Kode Gejala	Gejala
1.	D1	Asma
2.	D2	Bronkitis
3.	D3	Dyspnea
4.	D4	Pneumonia
5.	D5	PPOK
6.	D6	Tuberkulosis

3.3. Algoritma Naïve Bayes Classifier

Metode Naïve Bayes Classifier adalah pengklasifikasi probabilitas yang sederhana berdasarkan Teorema Bayes. Naïve Bayes Classifier merupakan algoritma pembelajaran yang terawasi(supervise), dan metode ini memiliki asumsi yang sangat kuat tentang independensi setiap kondisi atau peristiwa yang karena inilah metode ini disebut juga dengan metode naïf(naïve) [19].

Langkah langkah dalam proses perhitungan metode Naïve Bayes Classifier yaitu sebagai berikut [20]:

- a. Mencari nilai prior untuk tiap kelas dengan menghitung rata-rata kelas.

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

- b. Mencari nilai likelihood untuk tiap-tiap kelas.

$$L = \frac{F}{B} \quad (2)$$

- c. Mencari nilai posterior dari tiap kelas.

$$P(c) \prod P(a|c) \quad (3)$$

- d. Mencari nilai confidence dari tiap kelas.

$$Confidence(c) = \frac{P(c)}{\sum P(c)} * 100 \quad (4)$$

Sebelum melakukan proses perhitungan, data rekam medis pasien yang diperoleh akan dibagi menjadi 2 data, yaitu data training sebanyak 100 dan data testing sebanyak 34. Data yang telah melalui tahap pre-processing dan dipisah akan ditampilkan pada tabel 4.

Tabel 4. Data Penyakit

Keterangan	Total Data	Data Training	Data Testing
Asma	24	17	7
Bronkitis	11	8	3
Dyspnea	23	18	5
Pneumonia	26	19	7
PPOK	31	23	8
Tuberkulosis	19	15	4

Terdapat 6 klasifikasi yang terbentuk pada kolom output, yaitu Asma, Bronkitis, Dyspnea, Pneumonia, PPOK, dan Tuberkulosis. Untuk menentukan seorang pasien penyakit paru-paru mengidap salah satu penyakit tersebut ditentukan oleh beberapa parameter yang terdapat pada tabel 2 yang berisi gejala penyakit.

Berikut dijelaskan proses tahapan perhitungan dengan menggunakan metode Naïve Bayes Classifier dengan data contoh sebagai berikut :

Tabel 5. Contoh Uji Data

No	Atribut	Input Gejala
1.	Gejala Sesak Nafas	1

2.	Gejala Batuk	1
3.	Gejala Batuk Berdarah	0
4.	Gejala Berdahak	1
5.	Gejala Demam	1
6.	Gejala Lemas	0
7.	Gejala Nafsu Makan Berkurang	0
8.	Gejala Mual	1
9.	Gejala Muntah	0
10.	Gejala BAB	0
11.	Gejala BAK	0
12.	Gejala Riwayat Penyakit Asma	1
13.	Gejala Riwayat Penyakit Stroke	0
14.	Gejala Riwayat Penyakit Tuberkulosis	0
15.	Gejala Sakit Kepala	1
16.	Gejala Nyeri Ulu Hati	0
17.	Gejala Nyeri Dada	0
18.	Gejala Berat Badan Menurun	0
19.	Gejala Keringat Malam	0

Berdasarkan contoh data uji yang telah dipaparkan di tabel 5, proses perhitungan manual pada salah satu diagnosa dengan menggunakan algoritma Naïve Bayes Classifier adalah sebagai berikut :

Penyelesaian:

$$\frac{\text{Peluang Asma (Asma|Ya)}}{\text{Total Jumlah Penyakit Asma}} = \frac{17}{100} = 0,170$$

$$\text{G1 Peluang Sesak Nafas (Ya|Sesak Nafas)} = \frac{P(\text{Jumlah Sesak Nafas Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G2 Peluang Batuk (Ya|Batuk)} = \frac{P(\text{Jumlah Batuk Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G3 Peluang Batuk Berdarah (Tidak|Batuk Berdarah)} = \frac{P(\text{Jumlah Batuk Berdarah Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G4 Peluang Berdahak (Ya|Berdahak)} = \frac{P(\text{Jumlah Berdahak Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{10}{17} = 0,588$$

$$\text{G5 Peluang Demam (Ya|Demam)} = \frac{P(\text{Jumlah Berdahak Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{8}{17} = 0,470$$

$$\text{G6 Peluang Lemas (Tidak|Lemas)} = \frac{P(\text{Jumlah Lemas Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{15}{17} = 0,882$$

$$\text{G7 Peluang Nafsu Makan Berkurang (Tidak|Nafsu Makan Berkurang)} = \frac{P(\text{Jumlah Nafsu Makan Berkurang Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{15}{17} = 0,882$$

$$\text{G8 Peluang Mual (Ya|Mual)} = \frac{P(\text{Jumlah Mual Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{6}{17} = 0,352$$

$$\text{G9 Peluang Muntah (Tidak|Muntah)} = \frac{P(\text{Jumlah Muntah Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G10 Peluang BAB (Tidak|BAB)} = \frac{P(\text{Jumlah BAB Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G11 Peluang BAK (Tidak|BAK)} = \frac{P(\text{Jumlah BAK Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G12 Peluang Riwayat Penyakit Asma (Ya|Riwayat Penyakit Asma)} = \frac{P(\text{Jumlah Riwayat Penyakit Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{9}{17} = 0,529$$

$$\text{G13 Peluang Riwayat Penyakit Stroke (Tidak|Riwayat Penyakit Stroke)} = \frac{P(\text{Jumlah Riwayat Penyakit Stroke Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G14 Peluang Riwayat Penyakit Tuberkulosis (Tidak|Riwayat Penyakit Tuberkulosis)} = \frac{P(\text{Jumlah Riwayat Penyakit Tuberkulosis Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$\text{G15 Peluang Sakit Kepala (Ya|Sakit Kepala)} = \frac{P(\text{Jumlah Sakit Kepala Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{3}{17} = 0,176$$

$$G16 \text{ Peluang Nyeri Ulu Hati (Tidak|Nyeri Ulu Hati)} = \frac{P(\text{Jumlah Nyeri Ulu Hati Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$G17 \text{ Peluang Nyeri Dada (Tidak|Nyeri Dada)} = \frac{P(\text{Jumlah Nyeri Dada Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{15}{17} = 0,882$$

$$G18 \text{ Peluang Berat Badan Menurun (Tidak|Berat Badan Menurun)} = \frac{P(\text{Jumlah Berat Badan Menurun Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

$$G19 \text{ Peluang Keringat Malam (Tidak|Keringat Malam)} = \frac{P(\text{Jumlah Keringat Malam Asma})}{\text{Jumlah Klasifikasi Asma}} = \frac{17}{17} = 1$$

Join Probability Distribution (Ya|Asma)

Hitung Asma(Asma|databaru)

$$= \text{Asma (Asma|Ya)} * \text{Sesak Nafas (Ya|Sesak Nafas)} * \text{Batuk (Ya|Batuk)} * \text{Batuk Berdarah (Ya|Batuk Berdarah)} * \text{Berdahak (Ya|Berdahak)} * \text{Demam (Ya|Demam)} * \text{Lemas (Ya|Lemas)} * \text{Nafsu Makan Berkurang (Ya|Nafsu Makan Berkurang)} * \text{Mual (Ya|Mual)} * \text{Muntah (Ya|Muntah)} * \text{BAB (Ya|BAB)} * \text{BAK (Ya|BAK)} * \text{Riwayat Penyakit Asma (Ya|Riwayat Penyakit Asma)} * \text{Riwayat Penyakit Stroke (Ya|Riwayat Penyakit Stroke)} * \text{Riwayat Penyakit Tuberkulosis (Ya|Riwayat Penyakit Tuberkulosis)} * \text{Sakit Kepala (Ya|Sakit Kepala)} * \text{Nyeri Ulu Hati (Ya|Nyeri Ulu Hati)} * \text{Nyeri Dada (Ya|Nyeri Dada)} * \text{Berat Badan Menurun (Ya|Berat Badan Menurun)} * \text{Keringat Malam (Ya|Keringat Malam)}$$

$$= 0,170 * 1 * 1 * 1 * 0,588 * 0,470 * 0,882 * 0,882 * 0,352 * 1 * 1 * 1 * 0,529 * 1 * 1 * 0,176 * 1 * 0,882 * 1 * 1$$

$$= 0,00106595 \text{ atau } 0,11\%$$

Dari Hasil perhitungan, didapatkan nilai probabilitas yang sama dari setiap diagnosa penyakit yang ditampilkan pada tabel 6.

Tabel 6. Contoh Uji Data

Asma	Bronkitis	Dyspnea	Pneumonia	PPO K	Tuberkulosis
0,0010	0,0000	0,000	0,00006	0,0000	0,00000

6595	0000	02540	353	1296	00
------	------	-------	-----	------	----

Nilai probabilitas yang telah didapatkan dari setiap diagnosa penyakit, dilakukan perhitungan untuk mencari nilai confidence.

$$= 0,00106595 + 0 + 0,00002540 + 0,00006353 + 0,00001296 + 0$$

$$= 0,00116784$$

Setelah menjumlahkan semua nilai dari probabilitas, maka nilai confidence dapat dihitung. Berikut adalah perhitungan mencari nilai confidence :

Hitung Confidence Asma

$$\frac{0,00106595}{0,00116784} = 0,9128$$

Dilakukan perhitungan yang sama dari setiap nilai probabilitas dari setiap diagnose penyakit. Dari setiap diagnosa dan nilai tertinggi dari setiap confidence menjadi kemungkinan klasifikasi dari penerapan algoritma Naïve Bayes Classifier. Nilai confidence dapat dilihat dari tabel 7.

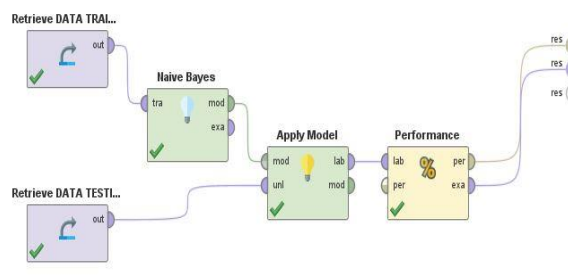
Tabel 7. Contoh Uji Data

Asma	Bronkitis	Dyspnea	Pneumonia	PPO K	Tuberkulosis
0,9128	0,0000	0,0217	0,0544	0,0111	0,0000

3.4. Klasifikasi Menggunakan Rapid Miner

Berikut adalah hasil proses dari Rapid Miner. Pada gambar 2 terdapat beberapa komponen yang digunakan yaitu Retrieve Data, Naïve Bayes, Apply Model, dan Performance. Retrieve berfungsi sebagai input data dalam bentuk excel, dipisah kedalam 2 bentuk yaitu Retrieve Data Training yang berfungsi sebagai data latih dan Retrieve Data Testing yang berfungsi sebagai data uji. Naïve Bayes berfungsi sebagai algoritma perhitungan yang digunakan dalam proses mining, terhubung dengan Apply Model yang menghubungkan Retrieve Data Testing dengan Naïve Bayes dan Retrieve Data Training. Performance berfungsi

untuk menampilkan hasil dari perhitungan menggunakan algoritma Naïve Bayes Classifier.



Gambar 2. Rangkaian Perhitungan Naïve Bayes Classifier

Untuk data yang digunakan yaitu 34 data testing dan 100 data training, terdapat 19 atribut yang merupakan gejala dari diagnosa penyakit paru-paru, yaitu Sesak Nafas, Batuk, Batuk Berdarah, Berdahak, Demam, Lemas, Nafsu Makan Berkurang, Mual, Muntah, BAB, BAK, Riwayat Penyakit Asma, Riwayat Penyakit Stroke, Riwayat Penyakit Tuberkulosis, Sakit Kepala, Nyeri Ulu Hati, Nyeri Dada, Berat Badan Menurun, dan Keringat Malam.

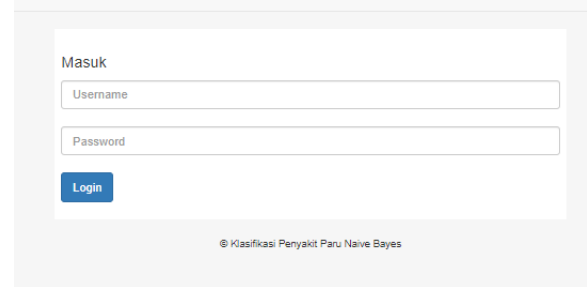
Diagnosa	Sesak Nafas	Batuk	Batuk Berdar...	Berdahak	Demam	Lemas
Dyspnea	0	1	0	1	1	0
Tuberkulosis	1	1	1	0	0	0
Asma	1	1	0	0	1	0
Bronkitis	1	1	0	0	0	0
Tuberkulosis	1	1	1	0	0	1
Tuberkulosis	1	0	0	0	1	1
PPOK	1	1	0	0	0	0
Asma	1	1	0	1	0	0
Pneumonia	1	1	0	0	1	0
PPOK	1	0	0	1	1	0

Gambar 3. Atribut Data

3.5. Rancangan Sistem

a. Halaman Login

Pada Halaman Login, ditampilkan portal yang mengarahkan pengguna untuk login saat menggunakan sistem yang dapat dilihat pada gambar 4.



Gambar 4. Halaman Login

b. Halaman Utama

Pada Halaman Utama, ditampilkan bagian utama dari sistem yang mana terdapat beberapa menu seperti Halaman Utama, Data Testing, Data Training, Proses Naïve Bayes, Klasifikasi dan Logout yang dapat dilihat pada gambar 5.



Gambar 5. Halaman Utama

c. Halaman Data Testing

Pada Halaman Data Testing, ditampilkan sekumpulan data yang berfungsi sebagai data latih dalam proses perhitungan yang dapat dilihat pada gambar 6.

Data Testing

No	Nama	g1	g2	g3	g4	g5	g6	g7	g8	g9	g10	g11	g12	g13	g14	g15	g16	g17	g18	g19	Di	
1	Abdul	1	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	PF
2	Andrian	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	Dy
3	Ani	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Dy
4	Ayu	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	Pn
5	Azizah	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	Dy
6	Daimar	1	1	0	1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	Pn
7	Dita	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	As
8	Dimas	1	1	0	1	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	Pn

Gambar 6. Halaman Data Testing

d. Halaman Data Training

Pada Halaman Data Training, ditampilkan sekumpulan data yang

berfungsi sebagai data uji dalam proses perhitungan yang dapat dilihat pada gambar 7.

Data Training

No	Nama	g1	g2	g3	g4	g5	g6	g7	g8	g9	g10	g11	g12	g13	g14	g15	g16	g17	g18	g19
1	Adi	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0
2	Ahmad	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1
3	Akbar	1	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0
4	Andika	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Anggi	1	1	1	0	0	1	1	0	0	0	0	0	0	1	0	0	0	1	1
6	Amwar	1	0	0	0	1	1	1	0	0	1	0	0	1	0	0	0	0	0	0
7	Anief	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Astuti	1	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0

Gambar 7. Halaman Data Training

- e. Halaman Proses Naïve Bayes Pada Halaman Proses Naïve Bayes, ditampilkan seluruh proses perhitungan, mulai dari mencari prior, likelihood, posterior, dan posterior pada data training yang dapat dilihat pada gambar 8.

Naive Bayes

1. Mencari Nilai Prior Kelas / Diagnosa (Data Training)

No	Kode	Diagnosa	Jumlah	Prior
1	D1	Asma	17/100	0.17
2	D2	Bronkitis	8/100	0.08
3	D3	Dyspnea	10/100	0.10
4	D4	Pneumonia	19/100	0.19
5	D5	PPOK	23/100	0.23
6	D6	Tuberkulosis	15/100	0.15
Total				1

2. Likelihood tiap-tiap kelas (Data Training)

Dataset training

Nama	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	G16	G17	G18	G19	Diagnosa
Adi	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	1	0	Dyspnea
Ahmad	1	1	1	0	0	0	1	0	0	1	0	0	0	0	0	0	0	1	1	Tuberkulosis

Gambar 8. Halaman Proses Naïve Bayes

- f. Halaman Klasifikasi Data Baru Pada Halaman Klasifikasi Data Baru, ditampilkan sebuah form mandiri yang dapat digunakan untuk melakukan proses klasifikasi terhadap data yang di input yang dapat dilihat pada gambar 9.

Klasifikasi Naive Bayes

Nama

Gejala yang dialami:

- Sesak Nafas
- Batuk
- Batuk Berdarah
- Berdahak
- Demam
- Lemas
- Nafsu Makan Berkurang
- Mual
- Muntah

Gambar 9. Halaman Klasifikasi Data Baru

3.6. Pengujian

- a. Menggunakan Data Testing Proses pengujian pertama yaitu dengan menggunakan data testing yang berjumlah 34 data. Tujuan dilakukan perhitungan dengan data testing yaitu untuk menghitung keakuratan dengan menggunakan metode Naïve Bayes Classifier. Hasil dari perhitungan dengan menggunakan sistem dapat dilihat pada gambar 10.

Nama	D1	D2	D3	D4	D5	D6	Diagnosa (aktual)	Diagnosa (prediksi)	Hasil
Abdul	0.000943	0.000007	0.003728	0.000099	0.536889	0.458335	PPOK	PPOK	Benar
Andrian	0.001211	0.235845	0.533113	0.000001	0.148075	0.001756	Dyspnea	Dyspnea	Benar
Ani	0.185217	0.119531	0.567404	0.000000	0.127581	0.000267	Dyspnea	Dyspnea	Benar
Ayu	0.008230	0.000072	0.009565	0.921690	0.060430	0.000013	Pneumonia	Pneumonia	Benar
Azizah	0.323215	0.219019	0.297047	0.000004	0.160715	0.000000	Dyspnea	Asma	Salah
Dainar	0.000463	0.000002	0.000593	0.924378	0.074563	0.000001	Pneumonia	Pneumonia	Benar
Dila	0.516974	0.087579	0.158733	0.165309	0.071406	0.000000	Asma	Asma	Benar
Dimas	0.008230	0.000072	0.009565	0.921690	0.060430	0.000013	Pneumonia	Pneumonia	Benar
Dinda	0.008230	0.000072	0.009565	0.921690	0.060430	0.000013	Pneumonia	Pneumonia	Benar
Donny	0.001371	0.444600	0.230234	0.227838	0.095955	0.000003	Bronkitis	Bronkitis	Benar

Gambar 10. Pengujian Menggunakan Sistem

Selain melakukan uji coba dengan menggunakan program, juga dilakukan proses perhitungan dengan menggunakan Rapid Miner. Dari proses perhitungan algoritma Naïve Bayes Classifier pada Rapid Miner dengan menggunakan 34 data testing dan 100 data training dengan perbandingan data 7:3 didapatkan nilai akurasi sebesar 97,06%. Dari 34 data testing, terdapat 1 data testing yang menghasilkan klasifikasi yang salah. Hasil pengujian dapat dilihat pada gambar 11.

accuracy: 97.06%

	true PPOK	true Dyspnea	true Pneumonia
pred. PPOK	8	0	0
pred. Dyspnea	0	4	0
pred. Pneumonia	0	0	7
pred. Asma	0	1	0

Gambar 11. Confusion Table

- b. Menggunakan Data Segar(Baru) Proses perhitungan kedua yaitu dengan menggunakan data segar. Setelah mendapatkan keakuratan dengan menggunakan data testing, maka

selanjutnya bisa melakukan klasifikasi pada halaman klasifikasi data baru. Setelah mengisi form, maka sistem akan menampilkan gejala yang dialami yang mana dapat dilihat pada gambar 12.

Hasil Prediksi User
Gejala yang dialami

Gejala	Ya/Tidak
Sesak Nafas	Ya
Batuk	Ya
Batuk Berdarah	Tidak
Berdahak	Tidak
Demam	Tidak
Lemas	Ya
Nafsu Makan Berkurang	Tidak
Mual	Tidak
Muntah	Ya
BAB	Tidak
BAK	Tidak
Riwayat Penyakit Asma	Tidak
Riwayat Penyakit Stroke	Tidak
Riwayat Penyakit Tuberkulosis	Tidak
Sakit Kepala	Tidak
Nyeri Ulu Hati	Tidak
Nyeri Dada	Tidak
Berat Badan Menurun	Tidak
Keringat Malam	Tidak

Gambar 12. Gejala Pasien

Setelah memasukkan gejala yang dialami, maka sistem akan melakukan proses perhitungan dan akan menunjukkan persentase perkiraan penyakit paru-paru yang paling mendekati dari 6 jenis penyakit paru yang tersedia oleh sistem. Dari gejala yang dimasukkan didapatkan nilai confidence tertinggi dengan nilai 0,984 dengan jenis prediksi penyakit PPOK. Rangkuman dari perhitungan yang dilakukan oleh program dapat dilihat pada gambar 13.

Naive Bayes Classification

Jenis Penyakit	Posterior	Confidence
Asma	0.000000	0.000
Bronkitis	0.000000	0.000
Dyspnea	0.000000	0.001
Pneumonia	0.000000	0.000
PPOK	0.000431	0.984
Tuberkulosis	0.000007	0.015

Berdasarkan perhitungan menggunakan model naive bayes, nilai confidence tertinggi yang didapat ialah 0.984, sehingga didapatkan bahwa pasien memiliki kemungkinan 98.36% Penyakit paru dengan jenis PPOK dibandingkan dengan 6 jenis penyakit paru lain

Gambar 13. Output Klasifikasi

4. KESIMPULAN

Kesimpulan dari penelitian ini adalah data mining dari proses klasifikasi dengan menggunakan metode Naïve Bayes Classifier efektif dalam mengklasifikasi penyakit paru-paru dengan nilai akurasi sebesar 97,06%, serta dalam proses implementasinya dalam bentuk program yang tentunya dapat mempermudah dalam mengklasifikasi penyakit paru. Sistem dapat menghasilkan output dengan memasukkan data gejala dari pasien paru-paru baru, dengan hasil perhitungan yang dihasilkan dapat menunjukkan perkiraan penyakit paru-paru yang paling mendekati berdasarkan gejala yang dialami pasien dibandingkan dengan 6 jenis penyakit paru lainnya.

5. REFERENSI

- [1] I. Majid and M. Endang Dian Setioningsih, ST, MT, Sumber, SST, "Monitoring BPM, Suhu dan Respirasi Tampil PC via Bluetooth dan Pengiriman Data via SMS (Monitoring BPM dan Pengiriman Data via SMS)," *Semin. Tugas Akhir*, 2018.
- [2] A. S. Pratama, S. Safrizal, and J. Iriani, "Sistem Pakar Mendiagnosa Penyakit Gangguan Pernafasan Oleh Asap Rokok Menggunakan Metode Dempster Shafer," *It (Informatic Tech. J.*, vol. 9, no. 1, p. 79, 2021, doi: 10.22303/it.9.1.2021.79-88.
- [3] R. Wana Dwi Putra, E. Prasetyo, and U. Chotijah, "Klasifikasi Penyakit Paru-Paru Menggunakan Metode Naive Bayes Studi Kasus Rs.Pku Muhammadiyah Ujung Pangkah Gresik," no. 0561, p. 740186, 2020, [Online]. Available: <http://eprints.umg.ac.id/3669/>.
- [4] S. J. S. Tyas, M. Febianah, F. Solikhah, A. L. Kamil, and W. A. Arifin, "Analisis Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining Untuk Memprediksi Kelulusan," *J. Teknol. Inf. Dan Komun.*, vol. 8, no. 1, pp. 86–99, 2021, [Online]. Available:

- <https://jurnal.plb.ac.id/index.php/tematik/article/view/576/365>.
- [5] D. Larassati, A. Zaidah, and A. Sarika, "Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes," *J. Rekayasa Elektro Sriwij.*, vol. 2, no. 2, pp. 151–157, 2021, doi: 10.36706/jres.v2i2.29.
- [6] R. Nofitri and J. Eska, "Implementasi Data Mining Klasifikasi C4.5 Dalam Menentukan Kelayakan Pengambilan Kredit," *Semin. Nas. R.*, vol. 9986, no. September, pp. 307–310, 2018, [Online]. Available: <https://www.jurnal.stmikroyal.ac.id/index.php/senar/article/view/192/136>.
- [7] M. Iqbal *et al.*, "Implementasi Metode Certainty Factor dalam Sistem Pakar Diagnosis Penyakit Paru-Paru Berbasis Android," *Coding J. Komput. dan Apl.*, vol. 07, no. 03, pp. 155–164, 2019.
- [8] F. N. Azizah and D. Juniati, "Analisis Jenis Penyakit Paru-Paru Berdasarkan Chest X-Ray Menggunakan Metode Fuzzy C-Means," *MATHunesa J. Ilm. Mat.*, vol. 9, no. 2, pp. 322–331, 2021, doi: 10.26740/mathunesa.v9n2.p322-331.
- [9] A. S. Mulyani, "Pemanasan Global, Penyebab, Dampak dan Antisipasinya," *Artik. Pengabd. Masy.*, pp. 1–27, 2021, [Online]. Available: http://repository.uki.ac.id/4908/1/PEMA_NASANGLOBAL.pdf.
- [10] J. Abidin, F. Artauli Hasibuan, K. Kunci, P. Udara, and D. Gauss, "Pengaruh dampak pencemaran udara terhadap kesehatan untuk menambah pemahaman masyarakat awam tentang bahaya dari polusi udara," *Pros. SNFUR-4*, no. September, pp. 1–7, 2019, [Online]. Available: <https://snf.fmipa.unri.ac.id/wp-content/uploads/2019/09/18.-OFMI-3002.pdf>.
- [11] D. Kurniawan, F. Eka Febriansyah, and A. Amidya, "Implementasi Pengujian Theorema Bayes Pada Studi Kasus Pemilihan Smartphone Android," *J. Komputasi*, vol. 6, no. 2, pp. 11–14, 2018, doi: 10.23960/komputasi.v6i2.1744.
- [12] M. F. Rifai, H. Jatnika, and B. Valentino, "Penerapan Algoritma Naive Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *Petir*, vol. 12, no. 2, pp. 131–144, 2019, doi: 10.33322/petir.v12i2.471.
- [13] I. Purnamasari, "Analisa Klasifikasi Loyalitas Siswa Lembaga Pendidikan Tari dengan Metode Naive Bayes," *J. Ilm. Komputasi*, vol. 19, no. 1, pp. 59–68, 2020, doi: 10.32409/jikstik.19.1.157.
- [14] B. M. Sapata and D. Juniati, "Klasifikasi Penyakit Paru Berdasarkan Citra X-Ray Thorax Menggunakan Metode Fraktal Box Counting," *MATHunesa*, vol. 7, no. 3, pp. 228–235, 2019.
- [15] S. Suprianto, "Perbandingan Metode Naive Bayes Classifier Dan Holistic Lexicon Based Dalam Analisis Sentimen Angket Mahasiswa," *JSI J. Sist. Inf.*, vol. 11, no. 2, pp. 1763–1771, 2019, doi: 10.36706/jsi.v11i2.9140.
- [16] A. C. Adi, D. P. Lestari, F. S. S. Elsa, and Y. Sabui, "Online School Sentiment Analysis in Indonesia on Twitter Using The Naive Bayes Classifier and Rapid Miner Tools," *Int. J. Innov. Sci. Res. Technol.*, vol. 71, no. 1, pp. 785–788, 2022, [Online]. Available: [https://ijisrt.com/assets/upload/files/IJISRT22JAN782_\(1\).pdf](https://ijisrt.com/assets/upload/files/IJISRT22JAN782_(1).pdf).
- [17] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, "Implementasi Data Mining dengan Algoritma Naive Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako," *J. Ris. Komput.*, vol. 8, no. 6, pp. 219–225, 2021, doi: 10.30865/jurikom.v8i6.3655.
- [18] D. Novianti, "Implementasi Algoritma Naive Bayes Pada Data Set Hepatitis

- Menggunakan Rapid Miner,” *Paradig. - J. Komput. dan Inform.*, vol. 21, no. 1, pp. 49–54, 2019, doi: 10.31294/p.v21i1.4979.
- [19] B. Gunawan, H. S. Pratiwi, and E. E. Pratama, “Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode Naive Bayes,” *J. Edukasi dan Penelit. Inform.*, vol. 4, no. 2, p. 113, 2018, doi: 10.26418/jp.v4i2.27526.
- [20] D. T. Satya and N. Hidayat, “Sistem Pakar Diagnosis Penyakit Sapi Ternak Potong Menggunakan Metode Naive Bayes - Certainty Factor,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3406–3410, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2626/978>.