



PENERAPAN METODE NAÏVE BAYES UNTUK KLASIFIKASI PENYAKIT ISPA

Yuwanda Purnamasari Pasrun

Program Studi Sistem Informasi, Universitas Sembilanbelas November Kolaka

e-mail: yuwanda@usn.ac.id

Abstrak

ISPA adalah infeksi yang menyerang saluran pernapasan manusia dibagi ke dalam dua jenis besar yaitu penyakit saluran pernapasan atas (ISPaA) atau bawah (ISPbA). ISPA biasanya menular, yang dapat menimbulkan berbagai spektrum penyakit yang berkisar dari penyakit tanpa gejala atau infeksi ringan sampai penyakit yang parah dan mematikan, tergantung pada patogen penyebabnya. ISPaA bila tidak diatasi dengan baik dapat berkembang menjadi ISPbA. Timbulnya gejala biasanya cepat, yaitu dalam waktu beberapa jam sampai beberapa hari. Penyebab ISPA dapat disebabkan kurangnya pengetahuan dan informasi terhadap ISPA serta keterbatasan penanganan jika terjadi pada balita yang dialami para orangtua. Agar proses pencegahan dapat dilakukan dengan tepat, maka perlu diketahui dulu klasifikasi jenis ISPA apa yang diderita. Klasifikasi merupakan proses memetakan data ke dalam satu atau beberapa kelas/label yang sudah didefinisikan sebelumnya. Klasifikasi dapat dilakukan dengan memanfaatkan data yang telah ada melalui proses *data mining*. Metode Naïve Bayes adalah salah satu metode dalam *data mining* yang dapat digunakan sebagai pengklasifikasian probalistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Berdasarkan pengujian, metode Naïve Bayes terbukti dapat menghasilkan akurasi yang maksimal yaitu 96% dibandingkan dengan penelitian sebelumnya menggunakan *5-fold cross validation* pada 50 data.

Kata kunci—ISPA, Data Mining, Klasifikasi, Naïve Bayes

Abstract

ISPA is an infection that attacks the human respiratory tract divided into two major types, namely upper respiratory tract disease (ISPaA) or lower (ISPbA). ISPA is usually contagious, which can cause a variety of diseases ranging from asymptomatic or mild infections to severe and deadly diseases, depending on the pathogen. ISPaA if not handled properly can develop into ISPbA. The onset of symptoms is usually rapid, which is within a few hours to several days. The cause of ISPA can be due to lack of knowledge and information about ISPA and limited handling if it occurs in infants experienced by parents. So that the prevention process can be carried out appropriately, it is necessary to know first what classification of ISPA suffers. Classification is the process of mapping data into one or several classes / labels that have been previously defined. Classification can be done by utilizing existing data through the data mining



process. Naïve Bayes method is one of the methods in data mining that can be used as a simple probabilistic classification that calculates a set of probabilities by adding up the frequency and combination of values from a given dataset. Based on testing, the Naïve Bayes method is proven to be able to produce maximum accuracy of 96% compared to previous studies using 5-fold cross validation on 50 data.

Keywords—ISPA, Data Mining, Classification, Naïve Bayes

1. PENDAHULUAN

Infeksi saluran pernapasan akut (ISPA) adalah infeksi yang menyerang saluran pernapasan manusia. ISPA adalah penyakit saluran pernapasan atas (ISPaA) atau bawah (ISPbA), biasanya menular, yang dapat menimbulkan berbagai spektrum penyakit yang berkisar dari penyakit tanpa gejala atau infeksi ringan sampai penyakit yang parah dan mematikan, tergantung pada patogen penyebabnya, faktor lingkungan, dan faktor pejamu [1]. ISPaA bila tidak diatasi dengan baik dapat berkembang menjadi ISPbA. Timbulnya gejala biasanya cepat, yaitu dalam waktu beberapa jam sampai beberapa hari. Gejalanya meliputi demam, batuk, dan sering juga nyeri tenggorok, *coryza* (pilek), sesak napas, *mengi*, atau kesulitan bernapas. Belakangan, kasus ISPA Atas adalah yang paling banyak terjadi sehingga diperlukan penanganan dengan baik karena dampak komplikasinya yang membahayakan adalah *otitis*, *sinusitis* dan *faringitis* [2].

Berdasarkan data dari Kementerian Kesehatan Republik Indonesia pada tahun 2016, ISPA merupakan penyakit yang bisa menyerang semua kalangan tanpa melihat umur dan wilayah. Di Indonesia, penyakit ISPA menjadi salah satu penyakit yang berbahaya di antaranya adalah pneumonia. Tahun 2014, angka cakupan penemuan pneumonia balita tidak mengalami perkembangan berarti, yaitu berkisar antara 20%-30%. Pada tahun 2015 terjadi peningkatan menjadi 63,45%. Angka kematian akibat pneumonia pada balita sebesar 0,16% mengalami peningkatan dibandingkan dengan tahun 2014 yang sebesar 0,08% [3].

Penyebab ISPA dapat disebabkan kurangnya pengetahuan dan informasi terhadap ISPA serta keterbatasan penanganan jika terjadi pada balita yang dialami para

orangtua. Oleh sebab itu perlu adanya tindakan dan pencegahan secara dini mengenai ISPA [4]. Agar proses pencegahan dapat dilakukan dengan tepat, maka perlu diketahui dulu klasifikasi jenis ISPA apa yang diderita.

Sebelumnya telah dilakukan penelitian untuk menentukan klasifikasi penyakit ISPA di antaranya adalah sistem pakar [1] dan metode klasifikasi dengan *data mining* pada penelitian [5] dan [6]. Secara internal, sistem pakar harus memperhatikan dua komponen utama yakni basis pengetahuan (*knowledge base*) yang diberikan oleh pakar dan mesin inferensi (*inference engine*) untuk menghasilkan kesimpulan sebagai respon terhadap kueri yang dihasilkan *user* [7]. Sedangkan klasifikasi dalam *data mining* merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia menggunakan data latih sebagai data pembangun model [8]. Dalam pembangunan model diperlukan suatu algoritma yang tepat.

Penelitian [5] menggunakan Algoritma *Decision Tree* (ID3), akan tetapi tidak menunjukkan pengujian akurasi sehingga belum bisa dipastikan keakuratan hasil klasifikasi. Sementara penelitian [6] menggunakan metode *Classification Based Association Rule* (CBA) dalam klasifikasi dengan *data mining* akan tetapi akurasi yang dihasilkan hanya sebesar 50%. Hal ini disebabkan dalam analisis asosiasi akan ada pola kombinasi gejala yang harus dibuang jika tidak termasuk dalam kriteria pola yang dicari (sering muncul). Untuk itu dibutuhkan suatu metode klasifikasi lain dalam *data mining* untuk meningkatkan akurasi menjadi lebih maksimal.

Metode Naïve Bayes merupakan salah satu metode dalam *data mining* yang dapat digunakan sebagai pengklasifikasian probalistik sederhana yang menghitung

sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari *dataset* yang diberikan. Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*training data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian [9]. Metode Naive Bayes telah digunakan dan terbukti handal dalam beberapa penelitian untuk klasifikasi penyakit seperti penyakit jantung [10], penyakit Tuberculosis [11], dan penyakit Kanker Serviks [12]. Oleh karena itu, berdasarkan data dan fakta yang telah disebutkan maka Metode Naive Bayes dapat digunakan untuk memperbaiki kinerja klasifikasi pada penelitian [6]. Penelitian ini mengusulkan implementasi metode Naive Bayes dalam klasifikasi penyakit ISPA menggunakan data puskesmas Lepo-Lepo agar hasil klasifikasi dapat lebih maksimal.

2. METODE PENELITIAN

2.1 Infeksi Saluran Pernapasan Akut (ISPA)

Infeksi saluran pernapasan akut berdasarkan wilayah infeksiya terbagi menjadi infeksi saluran pernapasan atas (ISPaA) dan infeksi saluran pernapasan bawah (ISPbA). ISPaA terdiri dari *Rhinitis*, *Sinusitis*, *Otitis Media* (Radang Telinga Tengah), *Stomatitis* (Mulut), *Faringitis* (Radang Tenggorokan). ISPbA terdiri dari infeksi *Epiglottis* (pita suara), *Bronchitis*, *Pneumonia*.

2.2 Data Mining

Data mining didefinisikan sebagai satu set teknik yang digunakan secara otomatis untuk mengeksplorasi secara menyeluruh dan membawa ke permukaan relasi-relasi yang kompleks pada set data yang sangat besar. Akan tetapi, teknik-teknik data mining dapat juga diaplikasikan pada representasi data yang lain, seperti domain data spasial, berbasis teks, dan multimedia (citra)[13]. Istilah *data mining* memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang dimiliki. *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, dan historis untuk

menemukan keteraturan, pola atau hubungan dalam *set* data berukuran besar [14].

2.3. Klasifikasi

Salah satu tugas *data mining* adalah klasifikasi data, yaitu memetakan (mengklasifikasikan) data ke dalam satu atau beberapa kelas/label yang sudah didefinisikan sebelumnya. Klasifikasi dapat didefinisikan secara detail sebagai suatu pekerjaan yang pelatihan/pembelajaran terhadap fungsi target f yang memetakan setiap vektor (set fitur) x ke dalam satu atau sejumlah label kelas y berbeda. Pekerjaan pelatihan tersebut akan menghasilkan suatu model yang akan disimpan sebagai memori [8].

Model dalam klasifikasi mempunyai arti yang sama dengan *blackbox*, di mana ada suatu model yang menerima masukan kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban sebagai keluaran dari hasil pemikirannya. Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan untuk memprediksi label kelas dari data baru yang belum diketahui label kelasnya. Dalam pembangunan model selama proses pelatihan tersebut diperlukan adanya suatu algoritma untuk membangunnya yang disebut sebagai algoritma pelatihan [15].

2.4. Naive Bayes

Naive Bayes merupakan salah satu metode *machine learning* yang menggunakan perhitungan probabilitas. Algoritma ini memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya [16]. Prediksi Bayes didasarkan pada teorema Bayes dengan formula umum yang ditunjukkan pada Persamaan (1).

$$P(H|A) = \frac{P(E|H) \times P(H)}{P(E)} \quad (1)$$

Keterangan :

$P(H|E)$ = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi

- $P(E/H)$ = Probabilitas sebuah bukti E terjadi akan mempengaruhi hipotesis H
- $P(H)$ = Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apa pun
- $P(E)$ = Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain

Kaitan antara naïve bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah bahwa hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi.

Jika X adalah vektor masukan yang berisi fitur dan Y adalah label kelas, Naïve Bayes dituliskan dengan $P(Y/X)$. Notasi tersebut berarti probabilitas kelas Y didapatkan setelah fitur-fitur X diamati. Notasi ini disebut juga probabilitas akhir (*posterior probability*) untuk Y sedangkan $P(Y)$ disebut probabilitas awal (*prior probability*) Y .

Selama proses pelatihan harus dilakukan pembelajaran probabilitas akhir $P(Y/X)$ pada model untuk setiap kombinasi X dan Y berdasarkan informasi yang didapat dari data latih. Dengan membangun model tersebut, suatu data uji X' dapat diklasifikasikan dengan mencari nilai Y' dengan memaksimalkan nilai $P(Y'/X')$ yang didapat. Persamaan (2) menunjukkan formulasi Naïve Bayes untuk klasifikasi.

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)} \quad (2)$$

Keterangan :

- $P(Y|X)$ = Probabilitas data dengan vektor X pada kelas Y . $P(Y)$ adalah probabilitas awal kelas Y
- $\prod_{i=1}^q P(X_i|Y)$ = Probabilitas independen kelas Y dari semua fitur dalam vektor X

Setiap set fitur $X = \{X_1, X_2, X_3, \dots, X_q\}$ terdiri dari q atribut (q dimensi). Nilai $P(X)$ selalu tetap sehingga dalam perhitungan prediksi nantinya tinggal dihitung bagian $P(Y) \prod_{i=1}^q P(X_i|Y)$ dengan memilih yang

terbesar sebagai kelas yang dipilih sebagai hasil klasifikasi.

2.5 Matriks Confusion

Matriks confusion merupakan tabel yang mencatat hasil kerja klasifikasi. Tabel 1 menunjukkan contoh matriks confusion yang melakukan klasifikasi masalah biner (dua kelas), misalnya kelas 0 dan 1.

Tabel 1 Matriks Confusion untuk Klasifikasi 2 Kelas

f_{ij}		Kelas hasil prediksi (j)	
		Kelas = 1	Kelas = 0
Kelas asli (i)	Kelas = 1	f_{11}	f_{10}
	Kelas = 0	f_{01}	f_{00}

Setiap sel f_{ij} dalam matriks menyatakan jumlah record/data dari kelas i yang hasil prediksinya masuk ke kelas j . Misalnya sel f_{11} adalah jumlah data dalam kelas 1 yang secara benar dipetakan ke kelas 1, dan f_{10} adalah data dalam kelas 1 yang dipetakan secara salah ke kelas 0 [15].

Berdasarkan isi matriks confusion, maka dapat diketahui jumlah data dari masing-masing kelas yang diklasifikasi secara benar yaitu ($f_{11} + f_{00}$) dan data yang diklasifikasikan secara salah yaitu ($f_{10} + f_{01}$). Kuantitas matriks confusion dapat diringkas menjadi dua nilai, yaitu akurasi dan laju *error*, disebut juga metric kinerja klasifikasi. Dengan mengetahui jumlah data yang diklasifikasikan dengan benar maka dapat diketahui akurasi hasil klasifikasi dan dengan mengetahui jumlah data yang diklasifikasikan dengan salah maka dapat diketahui laju *error* dari klasifikasi yang telah dilakukan.

$$Akurasi = \frac{jml \ data \ diklasifikasi \ benar}{jml \ klasifikasi \ yang \ dilakukan}$$

$$Akurasi = \frac{f_{11} + f_{00}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (3)$$

$$Laju \ Error = \frac{jml \ data \ diklasifikasi \ salah}{jml \ klasifikasi \ yang \ dilakukan}$$

$$Laju \ Error = \frac{f_{10} + f_{01}}{f_{11} + f_{10} + f_{01} + f_{00}} \quad (4)$$

2.5 K-Fold Coss Validation

K-Fold Cross Validation merupakan salah satu metode evaluasi klasifikator khusus dari cross validation. Prinsip metode Cross Validation adalah setiap data digunakan dalam jumlah yang sama untuk pelatihan dan dan tepat satu kali untuk pengujian. Misalnya set data dipecah menjadi dua bagian dengan ukuran yang sama. Pertama dipilih satu bagian pelatihan, sedangkan lainnya untuk pengujian. Selanjutnya adalah menukar peran, bagian yang tadinya menjadi set data latih sekarang ditukar menjadi set data uji, begitu pula sebaliknya. Pendekatan ini disebut dengan *two-fold cross validation*.

Total *error* didapatkan dengan menjumlahkan *error* yang didapat dari dua kali proses tersebut, sehingga setiap data berkesempatan satu kali menjadi data uji dan satu kali menjadi data latih [17]. *K-Fold Cross Validation* memecah set data menjadi *k* bagian set data dengan ukuran yang sama [15].

3. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Proses Klasifikasi

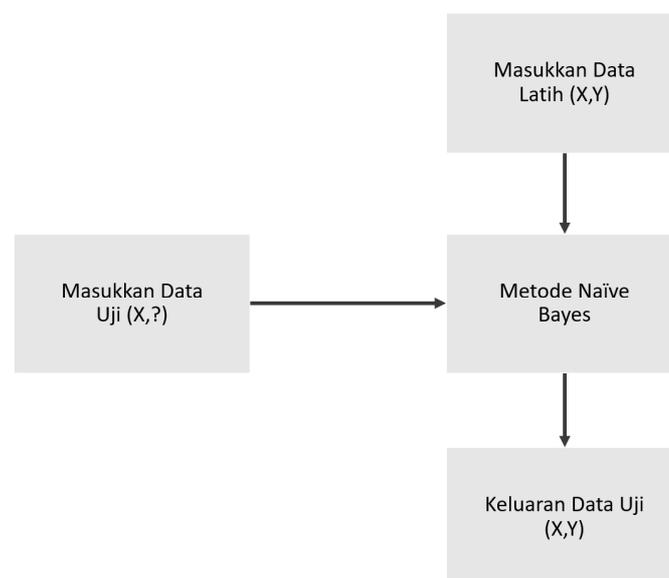
Penelitian ini menggunakan Metode Naïve Bayes untuk memperbaiki kinerja sistem klasifikasi penyakit ISPA pada penelitian [6]. Gambar 1 menunjukkan proses klasifikasi menggunakan metode Naïve Bayes.

Data yang digunakan sebanyak 50 data hasil diagnosis penyakit ISPA dari puskesmas Lepo-lepo Kendari. Data terdiri dari dua kelas yaitu kelas ISPaA sebanyak 25 data dan ISPbA sebanyak 25 data.

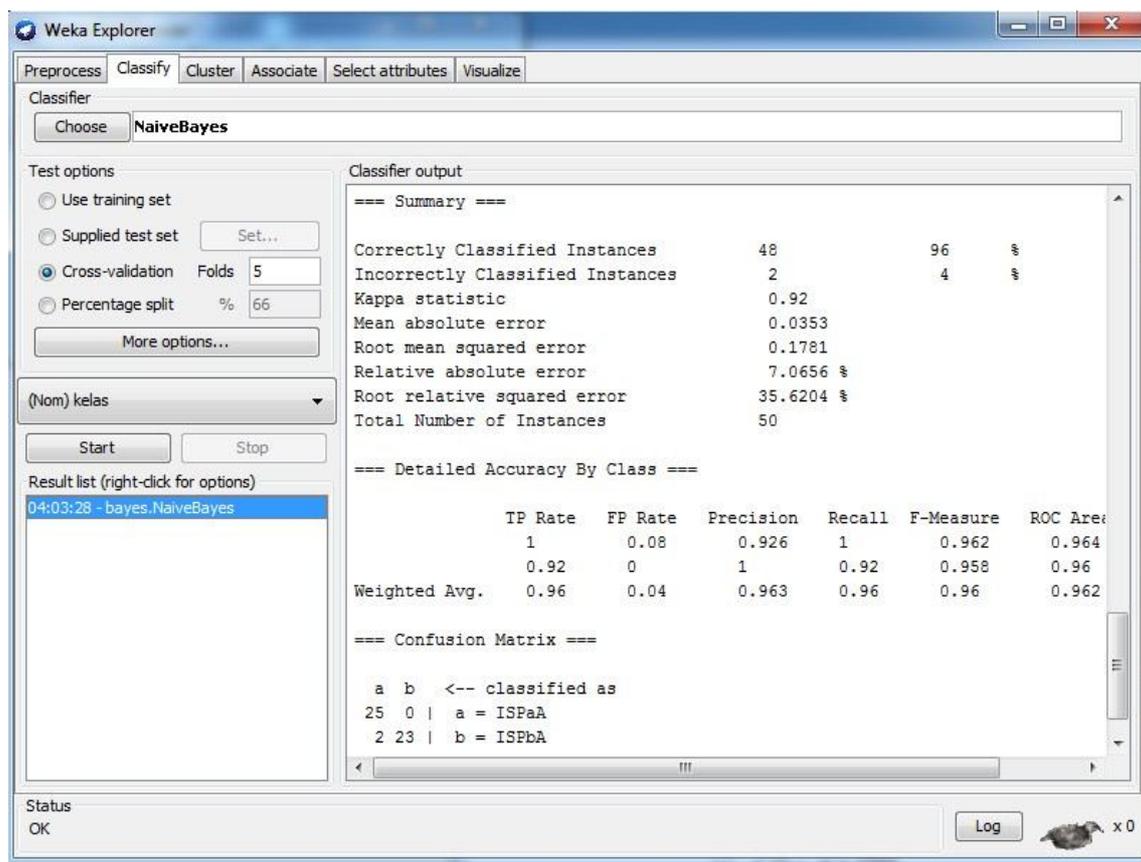
Pengujian metode Naïve Bayes sebagai klasifikator dilakukan untuk mengetahui apakah sistem berjalan sesuai dengan hasil yang diharapkan. Pengukuran kinerja menggunakan tools Weka untuk menghasilkan matriks confusion dan menghitung akurasi.

Untuk menguji akurasi Metode Naïve Bayes, data akan dibagi ke dalam dua bagian yaitu data latih dan data uji. Pembagian data latih dan data uji menggunakan metode pengujian *k-cross validation* dengan $k=5$, sehingga pengujian akan dilakukan sebanyak lima kali dan akurasi akhir merupakan akumulasi hasil rata-rata 5 akurasi dari 5 pengujian.

Gambar 2 menunjukkan hasil pengujian data menggunakan tools Weka dan Tabel 2 menunjukkan sampel data dari masing-masing kelas. Berdasarkan Gambar 2 dapat dilihat bahwa akurasi klasifikasi menggunakan *5-fold cross validation* sebesar 96%. Nilai ini diperoleh dari akurasi matriks confusion. Tabel 3 menunjukkan matriks confusion.



Gambar 1 Gambaran Umum Klasifikasi



Gambar 2 Hasil Implementasi Metode Naïve Bayes

Tabel 2 Sampel Data Puskesmas Lepo-lepo

NO	NAMA	USIA	BB	P/L	GEJALA	Kelas
1	rafa	1	10	L	Batuk(2 hari), Pilek(1 hari), Demam(3 hari)	ISPaA
2	akila	2	15	P	Pilek(1 hari), Sesak(1 hari)	ISPbA

Tabel 3 Matriks Confusion untuk Klasifikasi Penyakit ISPA Menggunakan Metode Naïve Bayes

f_{ij}		Kelas hasil prediksi (j)	
		ISPaA	ISPbA
Kelas asli (i)	ISPaA	25	0
	ISPbA	2	23

Dari matriks confusion didapatkan 48 dari 50 data uji yang diklasifikasi secara benar. Berdasarkan matriks confusion juga dapat dilihat terdapat 2 data kelas ISPbA diklasifikasi secara salah pada kelas kelas ISPaA. Akan tetapi nilai ini masih menunjukkan akurasi yang lebih maksimal dibandingkan dengan penelitian sebelumnya [6].

4. KESIMPULAN

Berdasarkan implementasi dan pengujian akurasi yang telah dilakukan maka dapat ditarik kesimpulan bahwa Metode Naïve Bayes mampu melakukan klasifikasi jenis penyakit ISPA dengan akurasi yang maksimal yakni 96%. Pengujian dilakukan menggunakan *5-fold cross validation* pada 50 data.

5. SARAN

Adapun saran untuk penelitian selanjutnya adalah perlunya penambahan kombinasi data untuk kelas ISPbA agar tidak terjadi kesalahan klasifikasi pada kelas atau label data lain serta penggunaan metode evaluasi klasifikator lain.

DAFTAR PUSTAKA

- [1] M. Marlina, W. Saputra, B. Mulyadi, B. Hayati, and J. Jaroji, "Aplikasi sistem pakar diagnosis penyakit ispa berbasis speech recognition menggunakan metode naive bayes classifier," *Digit. Zone J. Teknol. Inf. Dan Komun.*, vol. 8, pp. 58–70, Mar. 2017.
- [2] WHO, "Infection prevention and control of epidemic-and pandemic-prone acute respiratory diseases in health care." WHO Interim Guidelines, 2007.
- [3] H. T. Sihotang, E. Panggabean, and H. Zebua, "Sistem Pakar Mendiagnosa Penyakit Herpes Zoster Dengan Menggunakan Metode Teorema Bayes," INA-Rxiv, preprint, Aug. 2019.
- [4] Renita, "Penerapan Teorema Naive Bayes Pada Sistem Pakar Untuk Mendiagnosa Penyakit Infeksi Saluran Pernapasan Akut (ISPA) Pada Balita," *STMIK Nusa Mandiri*, pp. 1–75, 2016.
- [5] A. E. Pramadhani and T. Setiadi, "Penerapan Data Mining untuk Klasifikasi Prediksi Penyakit Ispa (Infeksi Saluran Pernapasan Akut) dengan Algoritma Decision Tree (Id3)," *Journal:eArticle*, Universitas Ahmad Dahlan, 2014.
- [6] H. Haryati, N. Ransi, and Y. P. Pasrun, "Penerapan Metode Cba (Classification Based On Assosiation Rule) Menggunakan Algoritma Apriori Untuk Klasifikasi Penyakit Ispa (Infeksi Saluran Pernapasan Akut)," *semanTIK*, Vol. 3, No. 2, Nov. 2017.
- [7] R. Rosnelly, *Sistem Pakar: Konsep dan Teori*. Andi, 2012.
- [8] E. Prasetyo, *Data Mining konsep dan Aplikasi menggunakan MATLAB*. Yogyakarta: Andi, 2012.
- [9] S. A. Pattekari and A. Parveen, "Prediction System For Heart Disease Using Naive Bayes," *Int. J. Adv. Comput. Math. Sci.*, vol. 3, no. 2, pp. 290–294, 2012.
- [10] N. A. Widiastuti, S. Santosa, and C. Supriyanto, "Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Particle Swarm Optimization Untuk Deteksi Penyakit Jantung," *Pseudocode*, vol. 1, no. 1, pp. 11–14, Feb. 2014.
- [11] R. A. Saputra, "Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Penyakit Tuberculosis (Tb): Studi Kasus Puskesmas Karawang Sukabumi," p. 8, 2014.
- [12] T. Praningki and I. Budi, "Sistem Prediksi Penyakit Kanker Serviks Menggunakan CART, Naive Bayes, dan k-NN," *Creat. Inf. Technol. J.*, Vol. 4, No. 2, pp. 83–93, Jan. 2018.
- [13] V. S. Moertini, "DATA MINING SEBAGAI SOLUSI BISNIS," *INTEGRAL*, Vol. 7, No. 1, p. 13, 2002.
- [14] M. Ridwan, H. Suyono, and M. Sarosa, "Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier," *J. EECCIS*, Vol. 7, No. 1, pp. 59–64, 2013.
- [15] E. Prasetyo, *Data mining mengolah data menjadi informasi menggunakan MATLAB*. Yogyakarta: Andi Offset, 2014.
- [16] S. L. B. Ginting and R. P. Trinanda, "Teknik Data Mining Menggunakan Metode Bayes Classifier Untuk Optimalisasi Pencarian Pada Aplikasi Perpustakaan (Studi Kasus: Perpustakaan Universitas Pasundan – Bandung)," *J. Teknol. Dan Inf. JATI*, Vol. 3, No. 2, pp. 37–50, Sep. 2013.
- [17] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*. New York: Pearson Education, 2006.

