

IMPLEMENTASI METODE NAÏVE BAYES UNTUK PREDIKSI DAERAH TERJANGKIT DEMAM BERDARAH DENGUE DI PROVINSI JAMBI

Daka Waru¹⁾, Reny Wahyuning Astuti²⁾

¹²³Universitas Nurdin Hamzah

Jl. Kol. Abunjani, Sipin, Kota Jambi

E-mail: warudaka12@gmail.com¹, r3ny4atuti@gmail.com²

ABSTRACT

DHF or Dengue Hemorrhagic Fever is a disease caused by one of the four dengue viruses, namely the dengue virus serotype-1, serotype-2, serotype-3, and serotype-4. Dengue fever is also a contagious disease. To be able to predict areas affected by dengue hemorrhagic fever (DHF), a method is needed that can be used to make it easier to anticipate the spread or control the spread of the disease. The purpose of this study was to predict the spread rate of dengue hemorrhagic fever (DHF) in Jambi Province. This implementation is assisted by using the Rapidminer application with data input on areas affected by dengue in Jambi Province. The process in this analysis is carried out with data that has been provided with the Naïve Bayes Method to predict areas affected by dengue in Jambi Province. The results of this study The highest accuracy value in the Rapidminer application is at 100%.

Keywords: *Data Mining, Dengue Fever, Jambi Province, Naive Bayes, Prediction.*

ABSTRAK

DBD atau Demam Berdarah Dengue adalah penyakit yang disebabkan oleh salah satu dari empat virus dengue, yaitu virus dengue serotype-1, serotype-2, serotype-3, dan serotype-4. Demam berdarah juga merupakan penyakit yang mudah menular. Untuk dapat memprediksi daerah terjangkit demam berdarah dengue (DBD) diperlukan metode yang dapat digunakan untuk mempermudah dalam mengantisipasi penyebaran ataupun penanggulangan penyebaran penyakit tersebut. Tujuan penelitian ini adalah untuk memprediksi tingkat penyebaran demam berdarah dengue (DBD) di Provinsi Jambi. Implementasi ini dibantu dengan menggunakan aplikasi Rapidminer dengan input data daerah terjangkit DBD di Provinsi Jambi. Proses dalam analisis ini dilakukan dengan data yang telah diberikan dengan Metode Naïve Bayes untuk memprediksi daerah terjangkit DBD di Provinsi Jambi. Hasil dari penelitian ini Nilai accuracy tertinggi pada aplikasi Rapidminer ada pada angka 100%.

Kata kunci: *Data Mining, Demam Berdarah, Provinsi Jambi, Naive Bayes, Prediksi.*

1. PENDAHULUAN

DBD atau Demam Berdarah Dengue adalah penyakit yang disebabkan oleh salah satu dari empat virus dengue, yaitu virus dengue serotype-1, serotype-2, serotype-3, dan serotype-4. Demam berdarah juga

merupakan penyakit yang mudah menular. Dimana penularannya dapat dipicu oleh faktor risiko tertentu, antara lain: pernah mengalami infeksi virus dengue sebelumnya, tinggal atau bepergian ke

daerah tropis, dan dapat dengan mudah menjangkiti bayi, anak-anak, orang lanjut usia, dan orang dengan kekebalan tubuh yang lemah. Di Provinsi Jambi, kejadian Demam Berdarah Dengue telah menyebar ke seluruh kabupaten/kota. Kota Jambi masih mencatat kasus tertinggi sepanjang tahun 2007 hingga tahun 2018, sesuai dengan *pattern of disease* dari penyakit DBD, yaitu *Urban Disease*. Hal ini dapat dimengerti mengingat Kota Jambi telah mempunyai fasilitas pelayanan kesehatan dengan laboratorium yang mendukung dan mobilitas penduduk dari dan ke daerah endemis DBD merupakan faktor risiko tingginya kasus DBD di Kota Jambi. Untuk dapat memprediksi daerah terjangkit demam berdarah dengue (DBD) diperlukan metode yang dapat digunakan untuk mempermudah dalam mengantisipasi penyebaran ataupun penanggulangan penyebaran penyakit tersebut.

Adapun metode yang digunakan adalah *Naïve Bayes*. Algoritma *Naïve Bayes* merupakan suatu bentuk klasifikasi data dengan menggunakan metode probabilitas dan statistik. Algoritma menggunakan teorema *Bayes* dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya.

Berdasarkan latar belakang tersebut maka peneliti tertarik untuk melakukan penelitian dalam bentuk laporan kerja praktek yang berjudul : Implementasi Metode *Naïve Bayes* Untuk Prediksi Daerah Terjangkit Demam Berdarah Dengue Di Provinsi Jambi.

2. METODOLOGI PENELITIAN

Penelitian ini dilaksanakan dengan menggunakan metode studi kasus (*cave research*). Studi kasus termasuk dalam penelitian analisis deskriptif yaitu penelitian yang dilakukan hanya membahas pada masalah tertentu untuk diamati dan dianalisis secara cermat. Penelitian ini terfokus pada satu objek tertentu dan terbatas dengan mempelajarinya sebagai suatu masalah (kasus), sehingga penelitian yang menggunakan studi kasus hasil penelitiannya tidak dapat digeneralisasikan, artinya penelitian hanya berlaku pada kasus yang dianalisis saja. Tetapi dalam prakteknya hasil penelitian dapat dijadikan sebagai contoh dari kasus selanjutnya.

2.1. Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan tahapan tahapan yang akan dilaksanakan dalam menyelesaikan masalah yang dibahas. Kerangka kerja penelitian ini berisi penjelasan tentang tahapan proses yang dilakukan selama melakukan kegiatan penelitian agar proses penelitian sesuai dengan tujuan yang diharapkan.

Adapun tahapan-tahapan kerangka kerja penelitian yang digunakan adalah sebagai berikut:



Gambar 1. Kerangka Kerja Penelitian

1. Mengidentifikasi Masalah

Merupakan salah satu langkah kerja penelitian yang boleh dikatakan paling penting, karena akan menentukan kualitas dari penelitian masalah penelitian bisa dideklarasikan sebagai pernyataan yang mempermasalahkan suatu variabel atau hubungan antara variabel pada suatu fenomena. Sedangkan variabel itu sendiri dapat didefinisikan sebagai pembeda antara satu dengan yang lain. Sedangkan sumber masalah dari suatu penelitian bisa muncul dari tiga hal, yaitu: *People And Problem, Program, dan Phenomenon*.

2. Pengumpulan Data

Pada tahap ini pengumpulan data ini, peneliti menggunakan metode wawancara, peneliti melakukan tanya jawab langsung dengan staf.

3. Mengimplementasi Perangkat Lunak

Tahap implementasi dilakukan dengan aplikasi *Microsoft Excel* dan *RapidMiner*. Aktivas implementasi adalah sekumpulan aktivitas dimana data yang telah di dapatkan akan masukkan kedalam aplikasi yang akan digunakan sehingga hasil dari proses yang dilakukan didalam aplikasi akan dapat digunakan nantinya.

4. Menganalisis kerja perangkat lunak

Tahap pengujian merupakan suatu teknik yang digunakan untuk menentukan bahwa implementasi yang dihasilkan telah mampu memecahkan masalah. Pada tahap ini implementasi yang baru dikerjakan diuji kemampuan dan keefektifannya sehingga didapatkan kekurangan dan kelemahan implementasi yang kemudian dilakukan pengkajian ulang dan perbaikan terhadap implementasi menjadi lebih baik dan sempurna.

5. Menarik Kesimpulan

Langkah selanjutnya dalam penelitian ini adalah mengambil kesimpulan. Dalam mengambil kesimpulan perlu kehati-hatian agar tidak terjerumus kepada kesimpulan yang terlalu sempit atau terlalu luas. Kesimpulan yang terlalu sempit bisa terjadi apabila ada ciri-ciri atau sifat atau fakta atau lainnya yang bersifat penting dalam ruang lingkup penelitian. Kesimpulan yang terlalu luas bisa terjadi apabila kesimpulan melebihi dari ruang lingkup penelitian. Suatu hal yang perlu diperhatikan adalah rumusan masalah atau pernyataan penelitian, harus terjawab didalam kesimpulan hasil penelitian.

6. Pembuatan Laporan

Pada tahap ini pembuatan laporan disusun berdasarkan hasil penelitian dengan menggunakan teknik pengumpulan data primer dan sekunder. Tujuannya adalah memberi gambaran secara utuh tentang data yang didapat dan hasil dari proses pemanfaatan data. Sehingga akan menghasilkan laporan penelitian yang baik dan sesuai dengan yang diharapkan.

2.2. Data Penelitian

Tabel 1. Data Training

Puskesmas	Kasus Meninggal	2016	2017
Memadai	Tidak	Rendah	Rendah
Memadai	Tidak	Sedang	Tinggi
Cukup	Tidak	Rendah	Rendah
Cukup	Tidak	Tinggi	Tinggi
Memadai	Tidak	Tinggi	Tinggi
Cukup	Ya	Tinggi	Tinggi
Cukup	Ya	Tinggi	Tinggi
Cukup	Tidak	Sedang	Sedang
Cukup	Tidak	Tinggi	Rendah
Cukup	Ya	Tinggi	Tinggi
Cukup	Tidak	Rendah	Tinggi

Tabel 2. Data Testing

Puskesmas	Kasus Meninggal	2016	2017
Cukup	Ya	Tinggi	Tinggi
Cukup	Ya	Tinggi	Tinggi
Cukup	Tidak	Sedang	Sedang
Cukup	Ya	Tinggi	Tinggi
Cukup	Tidak	Rendah	Tinggi
Memadai	Tidak	Rendah	Rendah
Memadai	Tidak	Sedang	Tinggi
Cukup	Tidak	Rendah	Rendah

2.3. Naïve Bayes

Naive Bayes merupakan sebuah pengklasifikasian probabilistik sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi dan kombinasi nilai dari dataset yang diberikan. Algoritma menggunakan teorema Bayes dan mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas. Definisi lain mengatakan Naive Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Naive Bayes didasarkan pada asumsi penyederhanaan bahwa nilai atribut secara kondisional saling bebas jika diberikan nilai output. Dengan kata lain, diberikan nilai output, probabilitas mengamati secara bersama adalah produk dari probabilitas individu (Manalu, E & Sianturi, FA & Manalu, MR 2017).

Algoritma Naïve Bayes Classifier memberikan sebuah kerangka pencocokan parsial tetapi masih belum maksimal. Untuk tercapainya hasil yang maksimal yaitu menerapkan bobot non-bilangan terhadap istilah indeks query dan dokumen. Biasanya nilai bobot dapat digunakan untuk menghitung tingkat kesamaan setiap dokumen yang tersimpan dalam sistem.

Hasilnya merupakan kumpulan dokumen yang sudah diambil jauh lebih akurat atau sesuai yang dibutuhkan dari pengguna. Ada dua langkah penting dalam klasifikasi dengan Naïve Bayes yaitu pelatihan dan pengklasifikasian :

a. Pelatihan pada tahap pelatihan ini pertama menghitung probabilitas setiap genre data pelatihan dengan rumus :

$$(C) = nc+s/N+sk \tag{1}$$

Yang kedua hitung katalog setiap nilai dari suatu atribut sebagai berikut:

$$(Ai = aj |c) = nj + s \tag{2}$$

Keterangan :

Nc = jumlah data

S = konstanta dari (s > 0)

K = jumlah katalog

P(Ai=aj|c) = nilai fitur aj-i untuk katalog

Vi = jumlah nilai dari atribut Ai

Nj = jumlah contoh dalam katalog

b. Pengklasifikasian pada tahap ini hitung probabilitas dari sebuah jenis untuk setiap katalog. Rumus :

$$(y|A) = (y) \prod P (Ai |y) \tag{3}$$

Keterangan :

P (y|A)= probabilitas jenis katalog y jika memiliki fitur A

P (y) = probabilitas dari sebuah jenis katalog y

P (Ai | y) = probabilitas yang memiliki nilai fiturAi dengan katalog y

Keuntungan menggunakan klasifikasi Naive Bayes adalah metode ini hanya memerlukan data training yang sedikit untuk membuat (mean dan varians dari variabel) yang diperlukan untuk klasifikasi. Karena variabel independen diasumsikan, hanya varians variabel masing-masing kelas yang harus ditentukan dan bukan keseluruhan matriks kovarians (Vijayarani & Dhayanand, 2015).

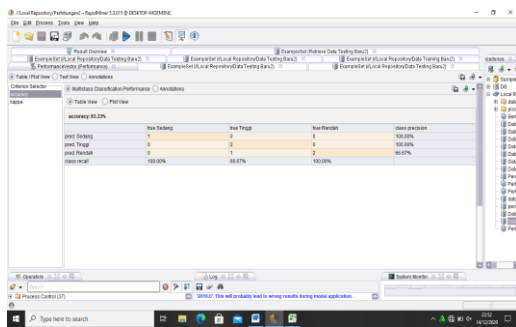
3. HASIL DAN PEMBAHASAN

Proses tahapan-tahapan implementasi ini akan menjelaskan bagaimana proses menggunakan aplikasi *Rapidminer* dengan metode *Naive Bayes* untuk mendapatkan prediksi daerah terjangkit DBD di Provinsi Jambi yang telah dilakukan peneliti. Adapun proses tahapannya adalah sebagai berikut:

3.1 Implementasi RapidMiner

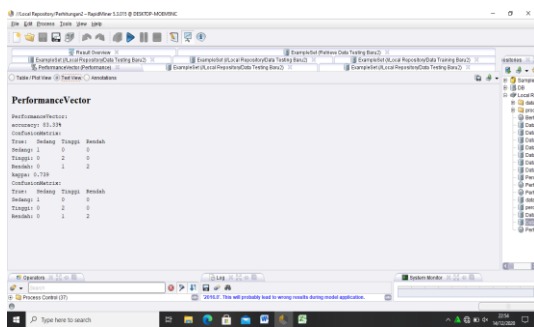
3.1.1 Tampilan Hasil Proses Naïve Bayes

Pada tampilan hasil proses ini menunjukkan hasil olahan dengan menggunakan metode *Naive Bayes*. Dapat dilihat pada gambar-gambar berikut ini :



Gambar 16. Tampilan *Table View*

Pada gambar di atas ditampilkan hasil perhitungan dari *Rapidminer* dalam bentuk tabel.



Gambar 17. Tampilan *Text View*

3.2 Implementasi Microsoft Excel

Untuk mengetahui ketepatan hasil dari perhitungan prediksi daerah terjangkit DBD di Provinsi Jambi di *Rapidminer*, peneliti melakukan perbandingan dengan melakukan perhitungan menggunakan *Microsoft Excel*, dengan hasil sebagai berikut :

3.2.1 Persiapan Data Training

Data training diperoleh dari data Dinas Kesehatan Provinsi Jambi, dari data diperoleh oleh peneliti, peneliti hanya menggunakan atribut yang diperlukan untuk perhitungan prediksi daerah terjangkit DBD di Provinsi Jambi, yaitu sebagai berikut :

	Puskesmas		Kasus Meninggal		2016	2017
	Training data	Memadai	Tidak	Rendah	Tinggi	Rendah
	Memadai	Tidak	Sedang	Tinggi	Tinggi	Tinggi
	Cukup	Tidak	Rendah	Tinggi	Rendah	Rendah
	Cukup	Tidak	Tinggi	Tinggi	Tinggi	Tinggi
	Memadai	Tidak	Tinggi	Tinggi	Tinggi	Tinggi
	Cukup	Ya	Tinggi	Tinggi	Tinggi	Tinggi
	Cukup	Ya	Tinggi	Tinggi	Tinggi	Tinggi
	Cukup	Tidak	Sedang	Sedang	Sedang	Sedang
	Cukup	Tidak	Tinggi	Rendah	Rendah	Rendah
	Cukup	Ya	Tinggi	Tinggi	Tinggi	Tinggi
	Cukup	Tidak	Rendah	Tinggi	Tinggi	Tinggi

Gambar 18. Data Training

3.2.2 Perhitungan Probabilitas

Perhitungan probabilitas dilakukan kepada semua kolom yang ada pada tabel, sehingga menghasilkan gambar-gambar berikut :

P(P= ...)	Rendah	Sedang	Tinggi
Kurang	0,00%	0,00%	0,00%
Cukup	66,67%	100,00%	71,43%
Memadai	33,33%	0,00%	28,57%
	100%	100%	100%

Gambar 19. Probabilitas Kolom Puskesmas

P(KM= ...	Rendah	Sedang	Tinggi
Ya	0,00%	0,00%	42,86%
Tidak	100,00%	100,00%	57,14%
	100%	100%	100%

Gambar 20. Probabilitas Kolom Kasus Meninggal

P(2016= ...	Rendah	Sedang	Tinggi
Rendah	66,67%	0,00%	14,29%
Sedang	0,00%	100,00%	14,29%
Tinggi	33,33%	0,00%	71,43%
	100%	100%	100%

Gambar 21. Probabilitas Kolom 2016

P(2017 Rendah)	P(2017 Sedang)	P(2017 Tinggi)
0,2727	0,0909	0,6364

3.2.3 Hasil Perhitungan Microsoft Excel

Pemeriksaan Kasus Meninggal	2016	2017	CLASS PREDICTION	Rendah	Sedang	Tinggi
Cukup	Tidak	Sedang	Sedang	0,0%	0,0%	3,7%
Cukup	Ya	Tinggi	Tinggi	0,0%	0,0%	13,0%
Cukup	Tidak	Rendah	Tinggi	72,1%	0,0%	3,7%
Memata	Tidak	Rendah	Rendah	6,1%	0,0%	1,5%
Memata	Tidak	Sedang	Tinggi	0,0%	0,0%	1,5%
Cukup	Tidak	Rendah	Rendah	72,1%	0,0%	3,7%

Gambar 23. Hasil Perhitungan Data Testing

Accuracy	Confusion Table	CLASS	Predicted	Rendah	Sedang	Tinggi
=	83%		Rendah	2	0	1
			Sedang	0	1	0
			Tinggi	0	0	2

Gambar 24. Hasil Prediksi dan Tingkat Akurasi

4. KESIMPULAN

Setelah peneliti melakukan analisis *data mining* daerah terjangkau DBD di Provinsi Jambi yang bertujuan untuk memprediksi daerah terjangkau DBD di Provinsi Jambi dengan metode *Naive Bayes*, peneliti menarik kesimpulan sebagai berikut :

1. Pengujian algoritma *Naive Bayes* dengan *Rapidminer* menunjukkan bahwa perhitungan manual dibandingkan dengan perhitungan aplikasi memiliki hasil yang sama.
2. Pengujian *accuracy* menunjukkan bahwa semakin banyak data maka tingkat akurasi semakin tinggi.

3. Nilai *accuracy* tertinggi pada aplikasi *Rapidminer* ada pada angka 100%.
4. Hasil penelitian ini diharapkan dapat digunakan oleh pengambil kebijakan untuk mempersiapkan tindakan pencegahan atau pembrantasan demam berdarah dengue.

DAFTAR PUSTAKA

- [1]. Manalu, E., Sianturi, F.A., & Manalu, M.R. Penerapan Algoritma Naïve Bayes Untuk Memprediksi Jumlah Produksi barang Berdasarkan Data Persediaan dan Jumlah Pemesanan pada CV. Papa dan Mama Pastries. *Jurnal Mantik Penusa*. 2017. 1(2).
- [2]. Vijayarani, S., & Dhayanand, S. Data Mining Classification Algorithms for Kidney Disease Prediction. *International Journal on Cybernetics & Informatics (IJCI)*. 2015. 4(4).
- [3]. Astuti, R.W., Puspitorini, S., & Akbar, F.K. Implementasi Teorema Naïve Bayes Pada Analisa dan Prediksi Bidang Pekerjaan Alumni Prodi Teknik Informatika STMIK Nurdin Hamzah Jambi. *FORTECH (Journal of Information Technology)*. 2020. Vol.4 No.1.