



PENERAPAN ALGORITMA NAÏVE BAYES UNTUK MEMPREDIKSI GEJALA DEMAM TIFOID PADA PUSKESMAS CIBADAK

Abdullah Khabari Kamil¹, Resti Yulistria², Apip Supiandi³, Gunawan⁴

¹Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika Sukabumi
Jl Cemerlang No. 8 Sukakarya Sukabumi, (0266) 6251992

abdullahskamil29@gmail.com

²Administrasi Bisnis, Fakultas Ekonomi dan Bisnis, Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 22, Jakarta Pusat

resti.res@bsi.ac.id

^{3,4}Ilmu Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika Sukabumi
Jl Cemerlang No. 8 Sukakarya Sukabumi, (0266) 6251992

apip.aup@bsi.ac.id, gunawan.gnz@bsi.ac.id

Abstrak-- Demam tifoid merupakan salah satu penyakit infeksi yang masih menjadi masalah kesehatan di negara berkembang, khususnya di Indonesia. Salmonella typhi merupakan bakteri penyebab demam tifoid yang dapat ditransmisikan melalui makanan maupun minuman terkontaminasi oleh fekes atau urin dari orang yang telah terinfeksi. Langkah pertama dalam pengelolaan penyakit demam tifoid adalah penetapan diagnosis yang tepat. Untuk mengurangi kesalahan deteksi dan menghindari keterlambatan diagnosis penderita demam tifoid, dapat dilakukan penerapan dan pemanfaatan teknik data mining. Salah satu algoritma yang dapat diterapkan adalah Naive Bayes Classifier, dengan diterapkannya algoritma Naive Bayes Classifier ini diharapkan para penderita dapat mengetahui kondisi kesehatannya dari demam tifoid yang mungkin saja terjadi, sehingga dapat langsung melakukan tindakan sebagai usaha untuk meminimalisir gejala yang terjadi dan diharapkan tindakan sejak dini ini membuat gejala lain yang akan terjadi justru tidak terjadi dan gejala yang ada berkurang. Naïve Bayes Classifier yang dikenal merupakan salah satu model klasifikasi yang baik dan sering digunakan. Hasil dari penelitian ini mendapatkan akurasi sebesar 93,71%. menggunakan rapid miner 5.2 dengan 142 dataset.

Kata kunci: Naïve Bayes Classifier, Data Mining, Demam Tifoid

Abstract - Typhoid fever is an infectious disease that is still a health problem in developing countries, especially in Indonesia. Salmonella typhi is a bacterium that causes typhoid fever which can be transmitted through food or drink contaminated by feces or urine from an infected person. The first step in managing typhoid fever is determining the right diagnosis. To reduce detection errors and avoid delays in diagnosis of typhoid fever sufferers, the application and utilization of data mining techniques can be carried out. One of the algorithms that can be applied is Naive Bayes Classifier, with the implementation of the Naive Bayes Classifier algorithm it is expected that sufferers can find out their health condition from typhoid fever that may occur, so they can immediately take action in an effort to minimize the symptoms that occur and are expected to take action early on. this makes other symptoms that will occur just do not occur and the symptoms are reduced. The Naïve Bayes Classifier is a well-known classification model and is often used. The results of this study get an accuracy of 93.71%. using rapid miner 5.2 with 142 datasets.

Keywords: Naive Bayes Classifier, Data Mining, Typhoid Fever

I. PENDAHULUAN

Demam tifoid merupakan salah satu penyakit infeksi yang masih menjadi masalah kesehatan di negara berkembang, khususnya di Indonesia.

Penyebarannya berkaitan dengan kepadatan penduduk, kesehatan lingkungan, sumber air dan sanitasi yang buruk, serta kebersihan pengolahan makanan yang masih rendah. Di Indonesia, insiden

* Korenpondensi.

Alamat E-mail : jurnal.larik@bsi.ac.id.

Diterima 30 July 2021; Direvisi 20 January 2021; Diterima 21 January 2021

© 2021 Jurnal Larik.

tifoid masih tergolong tinggi, bahkan menempati urutan ketiga di dunia [1].

Salmonella typhi merupakan bakteri penyebab demam tifoid yang dapat ditransmisikan melalui makanan maupun minuman terkontaminasi oleh feses atau urin dari orang yang telah terinfeksi. Diagnosis yang terlambat dan pengobatan yang salah dapat berakibat fatal pada penderita demam tifoid karena menyebabkan pendarahan saluran pencernaan yang mengakibatkan kematian. Metode deteksi gejala demam tifoid yang diinginkan mempunyai sensitifitas untuk mendeteksi satu set dalam sampel secara cepat, spesifik, dan akurat [2].

Untuk mengurangi kesalahan deteksi dan menghindari keterlambatan diagnosis demam tifoid, dapat dilakukan penerapan dan pemanfaatan teknik *data mining*. Pemanfaatan ini terjadi karena data mining bisa menemukan pola tersembunyi yang biasanya tidak ditemukan. Hal ini telah mendorong pengembangan sistem kecerdasan dan penunjang keputusan dalam bidang kesehatan untuk mendiagnosis keakuratan suatu penyakit, memprediksi tingkat keparahan berbagai jenis penyakit dan pengawasan kesehatan dari jauh. *Data mining* adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [3].

Teknik klasifikasi merupakan pendekatan fungsi klasifikasi dalam data mining yang digunakan untuk melakukan prediksi atas informasi yang belum diketahui sebelumnya. Beberapa algoritma yang dapat digunakan antara lain adalah algoritma *Decision Tree C.45*, *Artificial Neural Networks (ANN)*, *K-Nearest Neighbor (KNN)*, algoritma *Naive Bayes* serta algoritma lainnya [4].

Algoritma C4.5 memiliki kelebihan, yaitu mudah dimengerti, fleksibel, dan menarik karena dapat divisualisasikan dalam bentuk gambar (pohon keputusan)[5]. Kekurangannya yaitu, dapat terjadi *overlap* apabila hasil keputusan dan kriteria yang digunakan jumlahnya sangat banyak [6]. Kelebihan yang dimiliki *Artificial Neural Networks* adalah mengatasi permasalahan pelatihan klasifikasi dengan skala data yang luas dan *robust* terhadap *missing data*. Tetapi masih memiliki kekurangan dalam kecepatan *konvergen* yang buruk dan terjebak pada *local minimum* [7]. *Algoritma K-Nearest Neighbor (K-NN)* memiliki kelebihan karena sederhana, efektif dan telah banyak digunakan pada banyak masalah klasifikasi. Namun algoritma *K-NN*

memiliki kelemahan jika digunakan pada database yang besar karena membutuhkan waktu komputasi cukup tinggi [8].

Kelebihan algoritma *Naive Bayes Classifier* yaitu hanya membutuhkan jumlah data pelatihan yang kecil untuk menentukan estimasi parameter yang diperlukan dalam melakukan proses klasifikasi [9]. Dengan diterapkannya algoritma *Naive Bayes Classifier* ini diharapkan para penderita dapat mengetahui kondisi kesehatannya dari gejala demam tifoid yang mungkin saja terjadi, sehingga dapat langsung melakukan tindakan sebagai usaha untuk meminimalisir gejala yang terjadi dan diharapkan tindakan sejak dini ini membuat gejala lain yang akan terjadi justru tidak terjadi dan gejala yang ada berkurang.

II. METODE PENELITIAN

Tahapan penelitian mencakup beberapa langkah-langkah yang mengacu pada model CRISP-DM (*Cross Industry Standard Process for Data Mining*). Berikut ini rincian tahapan tersebut sebagai berikut:

1. Business Understanding

Tahapan ini melakukan penentuan tujuan penelitian yaitu penerapan algoritma *Naive Bayes Classifier* untuk memprediksi penyakit demam tifoid yang dilakukan di puskesmas Cibadak.

2. Data Understanding

Pada tahap ini yaitu mengumpulkan data sebagai bahan penelitian. Dalam hal ini dilakukan observasi lapangan untuk mendapatkan data pasien penyakit demam tifoid dari populasi sebanyak 142 data rekam medis dari jumlah populasi pasien tifoid selama tahun 2019 yang ada di Puskesmas Cibadak.

3. Data Preparation

Dalam tahap ini dilakukan perhitungan probabilitas terhadap 142 data rekam medis demam tifoid dengan *Naive Bayes* menggunakan *tools Rapid Miner 5.2*.

Rumus perhitungan yang digunakan pada algoritma *Naive Bayes* adalah sebagai berikut [10] :

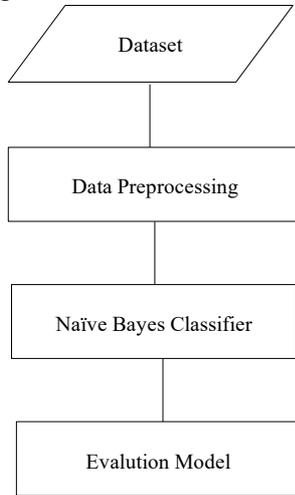
$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)} \quad (1)$$

Di mana:

- Y : Data dengan class yang belum diketahui
- X : Hipotesis data merupakan suatu class spesifik
- P(X|Y) : Probabilitas hipotesis X berdasar kondisi Y (posteriori probabilitas)
- P(X) : Probabilitas hipotesis X (prior probabilitas)
- P(Y|X) : Probabilitas Y berdasarkan kondisi pada

hipotesis X
 P(Y) : Probabilitas Y

4. Modelling



Gambar 1. Model yang diusulkan

Pada tahapan ini akan dilakukan proses pengujian data berdasarkan model yang diusulkan, tools yang digunakan adalah rapid miner 5.2.

5. Evaluation

Evaluasi dilakukan terhadap model yang dibangun dengan menggunakan *confusion matrix* untuk mengetahui nilai akurasi dan *area under curve* (AUC).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

6. Deployment

Tahap akhir dari penelitian ini menguji kembali model yang dibangun untuk mengetahui hasil evaluasi menggunakan data baru.

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data berjumlah 142 data dari jumlah keseluruhan pasien penderita penyakit demam tifoid. Data yang ada kemudian dihitung probabilitasnya lalu dicari nilai probabilitas dari setiap masing masing atribut untuk masing-masing Kelas $P(H)$ dan akan diimplementasikan pada *Rapid Miner*.

3.1. Menghitung Probabilitas Prior

Percobaan yang dilakukan dalam penelitian ini adalah dengan menghitung Probabilitas *Prior* serta Probabilitas *Posterior* menggunakan data

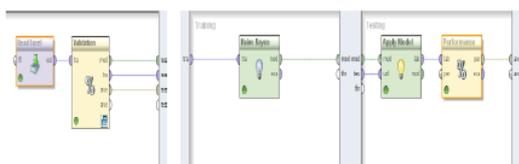
sebanyak 142 record data. Hasil Probabilitas *Prior* seperti pada tabel 1 terdapat dua kelas yang dibentuk yaitu kelas tifoid dan kelas tidak tifoid. Hasil perhitungan dapat dilihat pada tabel dibawah ini:

Tabel 1. Perhitungan Probabilitas Prior

ATRIBUT	JML DATA	KELAS TIFOID	KELAS TIDAK TIFOID	P(X Ci)	
				TIFOID	TIDAK TIFOID
TOTAL	142	85	57	0.5985 91549	0.401408 451
USIA	<=10	5	3	0.0352 94118	0.035087 719
	>50	23	18	0.2117 64706	0.087719 298
	11-15	23	13	0.1529 41176	0.175438 596
	16-20	23	14	0.1647 05882	0.157894 737
	21-25	19	9	0.1058 82353	0.175438 596
	26-30	44	26	0.3058 82353	0.315789 474
	31-35	1	1	0.0117 64706	0
JENIS KELAMIN	L	93	58	0.6823 52941	0.614035 088
	P	49	27	0.3176 47059	0.385964 912
DEMAM (HARI)	<5	36	0	0	0.631578 947
	>=6	106	85	1	0.368421 053
SAKIT KEPALA	YA	78	59	0.6941 17647	0.333333 333
	TIDAK	64	26	0.3058 82353	0.666666 667
PEGAL-PEGAL	KEJANG	17	17	0.2	0
	KERAM	63	42	0.4941 17647	0.368421 053
	KESEMUTAN	21	15	0.1764 70588	0.105263 158

	TIDAK K	41	11	30	0.1294 11765	0.526315 789
MUAL	YA	46	22	24	0.2588 23529	0.421052 632
	TIDAK K	96	63	33	0.7411 76471	0.578947 368
MUNTAH	YA	59	39	20	0.4588 23529	0.350877 193
	TIDAK K	83	46	37	0.5411 76471	0.649122 807
BATUK	YA	70	40	30	0.4705 88235	0.526315 789
	TIDAK K	72	45	27	0.5294 11765	0.473684 211
DENYUT NADI/MENI T	<80	34	0	34	0	0.596491 228
	80-100	108	85	23	1	0.403508 772
KEMBUNG	YA	101	80	21	0.9411 76471	0.368421 053
	TIDAK K	41	5	36	0.0588 23529	0.631578 947
DIARE	YA	85	50	35	0.5882 35294	0.614035 088
	TIDAK K	57	35	22	0.4117 64706	0.385964 912

Hasil hitungan probabilitas *prior* diatas dapat dibuatkan model *Naive Bayes classifier* dengan *Framework Rapidminer* versi 5.2.001 yaitu seperti gambar dibawah ini:



Gambar 2. Desain Model *Naive Bayes Classifier*

3.2. Menghitung Probabilitas *Posterior*

Probabilitas *Posterior* digunakan untuk menentukan kelas pada kasus baru, yang terlebih dahulu dihitung probabilitas *prior*nya. Jika ada kasus baru yang terlihat pada table dibawah ini:

Tabel 2. Perhitungan Probabilitas *Posterior*

Data X		P(X Ci)	
ATRIBUT	NILAI	TIFOID	TIDAK TIFOID
USIA	<=10	0.0352941 18	0.035087719

JENIS KELAMIN	L	0.6823529 41	0.614035088
DEMAM	>=6	1	0.368421053
SAKIT KEPALA	YA	0.6941176 47	0.333333333
PEGAL- PEGAL	KERAM	0.4941176 47	0.368421053
MUAL	YA	0.2588235 29	0.421052632
MUNTAH	YA	0.4588235 29	0.350877193
BATUK	TIDAK	0.5294117 65	0.473684211
DENYUT NADI/MENI T	80-100	1	0.403508772
KEMBUNG	YA	0.9411764 71	0.368421053
DIARE	YA	0.5882352 94	0.614035088

Setelah diketahui nilai probabilitas setiap atribut atau kriteria terhadap probabilitas tiap kelas atau $P(X|Ci)$, maka langkah selanjutnya adalah menghitung total keseluruhan probabilitas tiap kelas.

$$P(X|Ci) : P(X|hasil= TIFOID) \\ \Rightarrow 0.035294118 * 0.682352941 * 1 * 0.694117647 * 0.4 \\ 94117647 * 0.258823529 * 0.458823529 * 0.52941176 \\ 5 * 1 * 0.941176471 * 0.588235294 = \mathbf{0.00028750143}$$

$$P(X|hasil= TIDAK TIFOID) \\ \Rightarrow 0.035087719 * 0.614035088 * 0.368421053 * 0.333 \\ 333333 * 0.368421053 * 0.421052632 * 0.350877193 * \\ 0.473684211 * 0.403508772 * 0.368421053 * 0.61403 \\ 5088 = \mathbf{0.00000622711}$$

$$P(X|Pi) : P(X|hasil=TIFOID) P(TIFOID) \\ \Rightarrow 0.00028750143 * 0.598591549 = \mathbf{0.00017209592}$$

$$P(X|hasil= TIDAK TIFOID) P(TIDAK TIFOID) \\ \Rightarrow 0.00000622711 * 0.401408451 = \mathbf{0.00000249961}$$

Dari hasil perhitungan tersebut diketahui nilai $P(X|TIDAK TIFOID)$ lebih kecil dari pada nilai $P(X|TIFOID)$, sehingga dapat disimpulkan bahwa untuk kasus tersebut akan masuk kedalam klasifikasi TIFOID.

3.3. Evaluasi Model *Confusion Matrix*

Tabel 3. Confusion matrix *Naive Bayes*

Klasifikasi yang benar	Diklasifikasikan sebagai	
	Tifoid	Tidak Tifoid
Tifoid	63	7
Tidak Tifoid	2	50

Dengan mengetahui jumlah data yang diklasifikasi secara benar, maka dapat diketahui akurasi dari hasil *Rapid Miner* ini adalah sebesar

93,71%. Diketahui dari 142 data, 83 diklasifikasikan Tifoid sesuai dengan prediksi yang dilakukan dengan metode Algoritma *Naïve Bayes Classifier*, 7 data diprediksi Positif tetapi ternyata hasilnya Negatif, 50 data class Tidak Tifoid diprediksi sesuai, dan 2 data diprediksi Tidak Tifoid ternyata Tifoid.

3.4. AUC (Area Under Curve)

Kurva Receiver Operating Characteristic (ROC) digunakan untuk mengekspresikan data *confusion matrix*. Dari gambar berikut menerangkan terdapat garis horizontal yang mewakili nilai *false positif (FP)* dan garis vertikal mewakili nilai *true positif (TP)*. Dari gambar IV. dapat diketahui bahwa nilai *AUC (Area Under Curve)* model algoritma *naïve bayes classifier* adalah 0.979.



Gambar 3. AUC *Naive Bayes Classifier*

IV. KESIMPULAN DAN SARAN

Pengklasifikasian data mining terhadap gejala-gejala demam tifoid untuk memprediksi apakah gejala tersebut masih dalam kategori tifoid atau tidak tifoid, sehingga para penderita bisa mendeteksi sejak dini mengenai gejala ini dan diharapkan segera melakukan penanganan dengan cepat dan tepat. Teknik data mining berupa klasifikasi yang dilakukan pada data rekam medis demam tifoid yang diambil dari Puskesmas Cibadak yang dianalisis menggunakan algoritma *naïve bayes classifier* menghasilkan akurasi sebesar 93,71% dengan nilai *AUC excellent* yaitu 0.979.

Saran untuk penelitian selanjutnya, eksperimen menggunakan algoritma *Naive Bayes Classifier* menggunakan *dataset* lain untuk mengetahui performa *Naive Bayes Classifier* untuk *dataset* dengan jumlah data yang lebih beragam, atau bisa dilakukan eksperimen menggunakan data Demam Tifoid menggunakan metode algoritma data mining yang lain yang belum digunakan untuk *dataset* ini

untuk mengetahui algoritma lainnya dalam mengklasifikasikan gejala Demam tifoid.

REFERENSI

- [1] Purba, Y. P., & Ramadhian, M. R. (2019). *Efektivitas Antibakteri Ekstrak Etanol Mentimun (Cucumis sativus L .) terhadap Pertumbuhan Salmonella typhi Antibacterial Effectiveness of Cucumber (Cucumis sativus L .) Ethanol Extract on Salmonella typhi Growth*. 8, 144–149.
- [2] Hartono, D., Somantri, Iwan., Februanti, Sopia. (2019). Hipnosis Lima Jari dengan Pendekatan Spiritual Menurunkan Insomnia pada Lansia *Jurnal Kesehatan*, 187-192
- [3] Kustiyahningsih, Y., & Rahmanita, E. (2016). Aplikasi Sistem Pendukung Keputusan Menggunakan Algoritma C4.5. untuk Penjurusan SMA. *Jurnal Semantec*, 5(2), 101–108.
- [4] Wafiyah Fakihatun, Hidayat Nurul, P. R. S. (2017). Implementasi Algoritma Modified K-Nearest Neighbor (MKNN) untuk Klasifikasi Penyakit Demam. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 1(10), 1210–1219
- [5] Kumara, R., & Supriyanto, C. (2014). *Klasifikasi Data Mining Untuk Penerimaan Seleksi Calon Pegawai Negeri Sipil 2014 Menggunakan Algoritma Decision Tree C4.5*. 1–10.
- [6] Haryanto, F. F., & Hansun, S. (2017). Penerapan Algoritma C4.5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *Jurnal Teknik Informatika Dan Sistem Informasi*, 3(2), 95–103. Retrieved from <http://jurnal.mdp.ac.id/index.php/jatiasi/article/view/71>
- [7] Brian, T. (2016). *ANALISIS LEARNING RATES PADA ALGORITMA BACKPROPAGATION UNTUK KLASIFIKASI PENYAKIT DIABETES*. 3(1), 21–27.
- [8] Setiawan, T. A., Satria, R., Syukur, A., Komputer, F. I., & Nuswantoro, U. D. (2015). Integrasi Metode Sample Bootstrapping dan Weighted Principal Component Analysis untuk Meningkatkan Performa K Nearest Neighbor pada Dataset Besar. *Journal of Intelligent Systems*, 1(2), 76–81.
- [9] Sumpena, J., & H, N. K. (2019). *Analisis prediksi kelulusan siswa pkbm paket c dengan metoda algoritma naïve beyes*. 13(2), 127–133.
- [10] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," vol. 2, no. 3, pp. 207–217, 2015