

Penerapan Metode Naïve Bayes Dalam Pemeriksaan Kesehatan

Nurtriana Hidayati¹, Victor Gayuh Utomo², Nur Wakhidah³
Fakultas Teknologi Informasi dan Komunikasi
Universitas Semarang

Email : anna@usm.ac.id¹, victor@usm.ac.id², ida@usm.ac.id³

Abstract - Maintaining the safety and health of its workers by conducting daily health checks. The medical examination is carried out by a doctor or nurse at work. Workers who were declared unhealthy were not allowed to work on that day, in several cities, so that inspection standards were maintained, a web-based system was implemented that implemented a rule-based expert system. Naive Bayes algorithm which is included as an algorithm for classification based on data mining and machine learning activities. The initial process of Naive Bayes begins by studying existing data (complete with labels) through the training process. The results of this training are actually the probability of an outcome based on the variables used. Naive Bayes was chosen because it can study the results of past examinations and use them to make the next decision. This is felt to be a solution to the problems of previous system deficiencies that require the experience and expertise of health workers who are not necessarily equal.

Keywords: Worker Health, Prediction, Naive Bayes

Abstrak - Menjaga keselamatan dan kesehatan pekerjanya yaitu dengan cara melakukan pemeriksaan kesehatan harian. Pemeriksaan kesehatan dilakukan oleh dokter atau perawat di tempat kerja. Pekerja yang dinyatakan tidak sehat tidak boleh bekerja pada hari tersebut, di beberapa kota, agar standar pemeriksaan tetap terjaga, dibangun sistem berbasis web yang menerapkan *rule based expert system*. Algoritma Naive Bayes yang termasuk sebagai algoritma untuk melakukan klasifikasi yang berdasarkan atas kegiatan data mining dan machine learning. Proses awal Naive Bayes dimulai dengan mempelajari data yang ada (yang lengkap dengan labelnya) melalui proses training. Hasil training ini sebenarnya merupakan probabilitas dari suatu hasil berdasarkan variabel-variabel yang digunakan. Naive Bayes dipilih karena dapat mempelajari hasil pemeriksaan yang telah ada lalu dan menggunakannya untuk mengambil keputusan berikutnya. Hal ini dirasa dapat menjadi pemecahan masalah bagi kekurangan sistem sebelumnya yang membutuhkan pengalaman dan keahlian tenaga kesehatan yang belum tentu setara.

Kata Kunci : Kesehatan Pekerja, Prediksi, Naïve Bayes

I. PENDAHULUAN

Dalam menjalankan pekerjaannya, tenaga kerja wajib dijaga kesehatan dan keselamatannya. Baik tenaga kerja tersebut bekerja di darat, dalam tanah, di permukaan atau di lokasi lain. Keselamatan kerja diupayakan dengan menerapkan tindakan-tindakan pencegahan kecelakaan dan penyelenggaraan lokasi yang kerja yang aman. Hal tersebut termaktub pada UU no 1 tahun 1970 tentang Keselamatan Kerja. Perhatian pemerintah tentang Keselamatan dan Kesehatan Kerja (K3) ini juga tampak pada UU no 23 tahun 1992 tentang Kesehatan dan UU no 13 tahun 2003 tentang Ketenagakerjaan. Pemeriksaan kesehatan ini ternyata memiliki masalah dalam penerapannya. Wilayah kerja obyek

penelitian yang tersebar mengakibatkan pemeriksaan kesehatan ini dilakukan pada lokasi yang berbeda-beda. Proses pemeriksaan kesehatan ini tentunya membutuhkan tenaga kesehatan, yaitu dokter atau perawat. Pada lokasi tertentu, jumlah tenaga kesehatan ini dirasa kurang dibandingkan dengan tenaga kerja yang perlu diperiksa kesehatannya. Banyaknya beban kerja dan kelelahan dikhawatirkan dapat menimbulkan kesalahan pengambilan keputusan oleh tenaga kesehatan. Untuk mengatasinya, sebenarnya telah dimanfaatkan sistem pakar dengan metode rule-based. Hanya saja penggunaan metode rule-based ini masuh memiliki suatu kelemahan karena pada keadaan tertentu tenaga kesehatan dapat meng-

override hasil. Yang dimaksud dengan override adalah bahwa tenaga kesehatan dapat mengubah hasil keputusan sistem dari Layak menjadi Tidak Layak berdasarkan pertimbangan dari tenaga kesehatan yang memeriksa. Pertimbangan ini tidak dapat dimasukkan pada rule sistem karena memang bergantung pada pengalaman dan keahlian tenaga kesehatan. Hal ini dirasa tidak ideal karena pengalaman dan keahlian tenaga kesehatan dapat saja berbeda-beda [3].

Di sisi lain terdapat algoritma Naive Bayes yang termasuk sebagai algoritma untuk melakukan klasifikasi. Klasifikasi sendiri adalah dasar untuk data mining dan machine learning. Proses awal Naive Bayes dimulai dengan mempelajari data yang ada (yang lengkap dengan labelnya) melalui proses training. Hasil training ini sebenarnya merupakan probabilitas dari suatu hasil berdasarkan variabel-variabel yang digunakan. Berdasarkan probabilitas tersebut, Naive Bayes melakukan klasifikasi dilakukan dengan memberikan label pada data yang tidak memiliki label [4]. Pemanfaatannya pun cukup beragam, dari mengenali sentimen pelanggan [5] hingga pengolahan tulisan dalam bahasa Arab [6].

Pada penelitian ini diusulkan penambahan fitur untuk melengkapi sistem yang sudah ada dengan menggunakan Naive Bayes. Naive Bayes dipilih karena dapat mempelajari hasil pemeriksaan yang telah ada lalu dan menggunakannya untuk mengambil keputusan berikutnya. Hal ini dirasa dapat menjadi pemecahan masalah bagi kekurangan sistem sebelumnya yang membutuhkan pengalaman dan keahlian tenaga kesehatan yang belum tentu setara.

II. TINJAUAN PUSTAKA

a. Naive Bayes

Klasifikasi adalah masalah mendasar pada data mining dan machine learning. Algoritma melakukan proses belajar (learning) dengan menganalisa sekumpulan data, disebut data training, untuk membentuk fungsi klasifikasi.

Data training akan berisi sekumpulan variabel atau fitur dan label atau hasil [4].

Naïve Bayes termasuk salah satu metode klasifikasi yang memiliki kemampuan yang kompetitif dibandingkan dengan metode klasifikasi berdasarkan data yang sejenis seperti jaringan syaraf tiruan, support vector machine (SVM) dan k-nearest neighbor [7].

Naïve Bayes melakukan klasifikasi dengan asumsi bahwa setiap fitur saling bebas atau tidak mempengaruhi satu sama lain. Ini adalah asal kata naïf pada nama metode ini. Yang menarik adalah bahwa ketika asumsi ini tidak dipenuhi, klasifikasi dengan Naïve Bayes dapat berjalan dengan baik [8].

Metode Naïve Bayes menggunakan teori Bayes yang dapat dilihat pada rumus berikut

$$p(c|E) = \frac{p(E|c)p(c)}{p(E)}$$

Pada rumus tersebut p adalah fungsi kemungkinan, c menyatakan label atau hasil sedangkan E menyatakan kumpulan variabel atau fitur. Naïve Bayes dapat bekerja dengan berbagai jenis nilai fitur yang menyebabkan adanya beberapa implementasi seperti Multinomial Naïve Bayes, Bernoulli Naïve Bayes dan Gaussian Naïve Bayes. Pada penelitian ini digunakan Gaussian Naïve Bayes yang sesuai dengan nilai fitur yang kontinu. Rumus yang digunakan untuk melakukan prediksi adalah sebagai berikut

$$p(\vec{w}|c) = \prod_{v=1}^V \mathcal{N}(w_v | \mu_{c,v}, \sigma_{c,v}^2) = \prod_{v=1}^V \frac{1}{\sqrt{2\pi\sigma_{c,v}^2}} \exp\left(-\frac{(w_v - \mu_{c,v})^2}{2\sigma_{c,v}^2}\right)$$

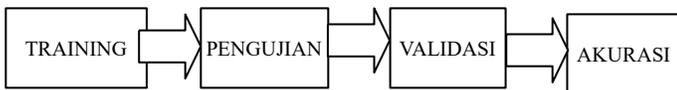
Dengan μ menyatakan mean atau rata-rata dan σ menyatakan varian [9].

III. METODOLOGI PENELITIAN

Terdapat sembilan variabel yang digunakan pada penelitian ini yaitu

1. Anamnesa
2. Pemeriksaan fisik
3. Tekanan darah sistole
4. Tekanan darah diastole
5. Nadi
6. Suhu
7. Respiration rate
8. Uji Romberg
9. Uji alkohol

Seperti pada penjelasan algoritma Naive Bayes maka proses penelitian ini akan berlangsung sebagai berikut



Penelitian dimulai dengan melakukan training terhadap data pemeriksaan kesehatan yang sudah ada dengan label Layak atau Tidak Layak. Pengujian kemudian dilakukan terhadap data pemeriksaan kesehatan tanpa label. Hasil pengujian kemudian divalidaasi terhadap nilai sesungguhnya. Hasilnya akan diperoleh ketepatan dari sistem yang dibangun.

IV. PEMBAHASAN

Algoritma Naïve Bayes pada penelitian ini diimplementasikan menggunakan Python dengan pustaka pemrograman NumPy dan SciKit. Tahap training pada penelitian ini dilakukan dengan menggunakan data pemeriksaan kesehatan yang sudah ada. Penelitian ini melakukan training sebanyak 3 kali. Setiap training, jumlah data yang digunakan bertambah untuk melihat efek perubahan training. Training pertama menggunakan data bulan Juli 2018 (1 bulan) sebanyak 5215 data. Training kedua dilakukan terhadap data bulan Juli hingga Agustus 2018 (2 bulan) sebanyak 20181 data. Training ketiga

menggunakan data bulan Juli-September 2018 (3 bulan) sebanyak 35413 data.

Pengujian untuk setiap hasil training dilakukan terhadap 4 dataset. Dataset pertama adalah data pemeriksaan kesehatan minggu pertama bulan November 2018 sebanyak 3446 data. Dataset kedua adalah data pemeriksaan kesehatan minggu kedua bulan November 2018 sebanyak 3052 data. Dataset ketiga adalah data pemeriksaan kesehatan minggu ketiga bulan November 2018 sebanyak 3142 data. Terakhir, dataset keempat adalah data pemeriksaan kesehatan minggu keempat bulan November 2018 sebanyak 3776 data.

Sistem yang digunakan pada penelitian ini akan melakukan validasi dengan membandingkan hasil prediksi kelayakan kesehatan (hasil pengujian dengan Naïve Bayes) dengan hasil kelayakan kesehatan yang ada. Rekap hasil validasi dapat dilihat pada tabel 1.

Tabel 1. Rekap Hasil Validasi Pengujian

Data Uji	Prediksi	Frekuensi		
		Training 1	Training 2	Training 3
Tidak Layak	Tidak Layak	10	8	8
Tidak Layak	Layak	8	10	10
Layak	Tidak Layak	38	26	26
Layak	Layak	13416	13372	13372

Dua kolom pertama pada tabel 1 menyatakan kombinasi nilai kelayakan kesehatan, kolom pertama berdasarkan data uji sedangkan pada kolom kedua berdasarkan prediksi (perhitungan dengan Naïve Bayes). Tiga kolom berikutnya menyatakan frekuensi hasil pengujian untuk semua dataset. Misalnya, pengujian dengan data training pertama, terdapat 10 prediksi Tidak Layak yang tepat. Contoh lain, pengujian dengan data training ketiga, terdapat 10 prediksi Layak yang tidak tepat. Akurasi didapatkan dengan melakukan pembagian antara total frekuensi prediksi yang tidak tepat dengan total pengujian. Hasil perhitungan akurasi pada penelitian ini dapat dilihat pada tabel 2.

Tabel 2. Akurasi Prediksi

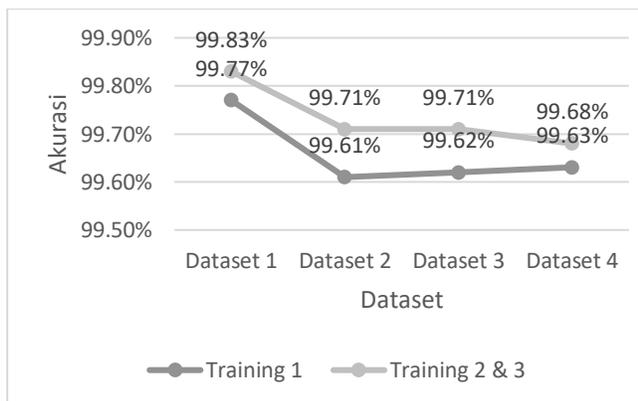
Trainin g	Dataset 1	Dataset 2	Dataset 3	Dataset 4
1	99.77 %	99.61 %	99.62 %	99.63 %
2	99.83 %	99.71 %	99.71 %	99.68 %
3	99.83 %	99.71 %	99.71 %	99.68 %

Kolom pertama pada tabel 2 menyatakan data training yang digunakan sedangkan keempat kolom berikutnya menyatakan dataset yang digunakan pada pengujian. Misalnya, akurasi pengujian dengan data training kedua terhadap dataset 3 adalah sebesar 99.71%.

Pembahasan

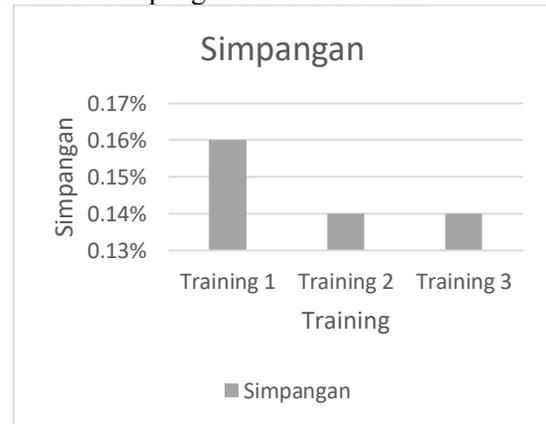
Pada grafik 1 dapat dilihat bahwa akurasi prediksi antara semua data training mendapat nilai tinggi (lebih dari 99%). Akurasi pengujian dengan menggunakan data training 2 dan 3 mendapatkan hasil yang sama maka ditampilkan secara berhimpit.

Grafik 1. Perbandingan Akurasi Prediksi



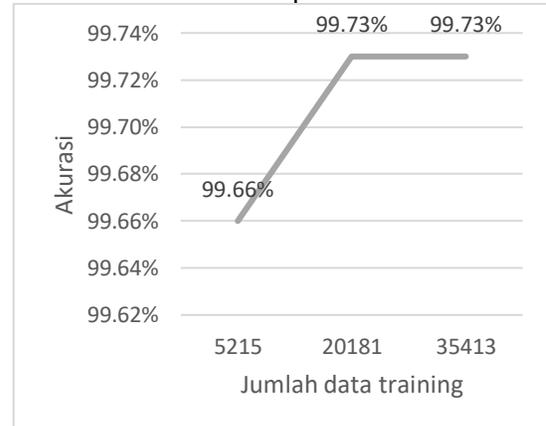
Akurasi terbaik pada grafik 1 dapat mencapai nilai 99.83% dengan akurasi terburuk pada 99.61%. Berikutnya, simpangan prediksi diperoleh dengan mengurangi nilai prediksi tertinggi dan nilai prediksi terendah dari masing-masing dataset untuk setiap data training. Hasil dapat dilihat pada Grafik 2.

Grafik 2. Simpangan Akurasi Prediksi



Pada Grafik 2 dapat dilihat bahwa hasil prediksi cenderung stabil. Selisih akurasi hanya berkisar pada angka 0.14% sampai 0.16%.

Grafik 3. Akurasi Terhadap Jumlah Data Training



Dengan mengolah jumlah data training dibandingkan dengan akurasi terdapat tren yang menarik yang dapat dilihat pada grafik 3. Penambahan data training dari training pertama dan training kedua meningkatkan akurasi hingga 0.07% atau jika dilihat pada tabel 1 mengurangi prediksi tidak tepat sebanyak 10. Sementara peningkatan jumlah data pada training kedua dan training ketiga, ternyata tidak memberikan pengaruh pada akurasi prediksi. Berdasarkan cara kerja Naïve Bayes, sebenarnya makin lengkap data training maka akurasi prediksi akan meningkat. Makin lengkap data membuat Naïve Bayes dapat mengenali lebih banyak kemungkinan variabel untuk menghasilkan label atau keputusan tertentu.

Hasil penelitian menunjukkan bahwa penambahan data training tidak secara langsung menambah keragaman data training. Akibatnya, akurasi tidak secara langsung meningkat walaupun jumlah data training bertambah.

V. KESIMPULAN

Pada penelitian ini digunakan metode Naïve Bayes untuk menentukan kelayakan pekerja berdasarkan hasil pemeriksaan kesehatannya. Metode Naïve Bayes digunakan sebagai perkembangan dari penelitian sebelumnya yang menggunakan metode Rule-Based. Terdapat 9 variabel yang digunakan pada penelitian ini yaitu, anamnesa, pemeriksaan fisik, tekanan darah systole, tekanan darah diastole, nadi, suhu, respiration rate, uji Romberg dan uji alkonol. Hasil penelitian menunjukkan bahwa metode Naïve Bayes mampu memberikan prediksi dengan akurasi yang baik. Akurasi maksimum yang dicapai pada penelitian ini adalah sebesar 99.83%. Hasil prediksi juga stabil dengan simpangan nilai maksimal 0.16%. Penelitian ini juga memperlihatkan, sampai batas tertentu, penambahan jumlah data training akan meningkatkan akurasi prediksi dengan metode Naïve Bayes.

SARAN

Tidak semua variable atau fitur pada penelitian ini memiliki nilai kontinu sehingga penelitian lebih lanjut untuk metode Naïve Bayes yang bukan Gaussian masih terbuka. Peningkatan akurasi dibandingkan dengan data training juga dapat diteliti lebih lanjut untuk melihat relasinya.

Kesalahan prediksi Layak cukup besar jika dibandingkan seluruh data uji Tidak Layak. Penelitian lebih lanjut diperlukan untuk mendapatkan perbaikan akurasi khusus untuk hal ini.

DAFTAR PUSTAKA

- [1] Priatna, Hendi dan Andika, Fauziah. 2018. Faktor yang Berhubungan dengan Kesadaran Keselamatan dan Kesehatan Kerja di Lingkungan Lanud Maimun Saleh Sabang. *Journal of Healthcare Technology and Medicine* Vol 4 No 1 Tahun 2018 Universitas Ubudiyah Indonesia
- [2] Saputro, Priyargo Bregas dan Riandadari, Dyah. 2019. Analisis Identifikasi Potensi Bahaya dalam Upaya Pencegahan Kecelakaan Kerja Metode Job Safety Analysis pada Proses Produksi di PT Infoglobal Teknologi Semesta. *JPTM* Vol 8 No 1 Tahun 2019
- [3] Utomo, Victor Gayuh dan Putra, Toni Wijanarko Adi. 2019. Sistem Pakar Penentuan Kelayakan Kesehatan Pekerja dengan Metode Rule-Based. Seminar Nasional Sains dan Teknologi 10, Semarang, Indonesia
- [4] Zhang, H. 2004. The Optimality of Naive Bayes. *Proceedings of the Seventeenth International Florida Artificial Intelligence Research Society Conference USA*
- [5] Irsyad, H., Farisi. A. dan Pribadi, Muhammad Rizky. 2019. Klasifikasi Opini Masyarakat Terhadap Jasa ISP MyRepublic dengan Naive Bayes. *JNTETI* Vol 8 No 1 Tahun 2019
- [6] Ahmed, Amjed Abbas. 2018. Intelligent Arabic Text Categorization: Initial Study and Proposed Methodology on Classifying Arabic Text Using Enhanced Naive Bayes Classification Approach. *Journal of Advanced Research in Dynamical and Control Systems* Vol 10 No 10 Tahun 2018
- [7] Informasi et al., “NAIVE BAYES UNTUK ANALISIS SENTIMEN,” vol. 11, no. April, pp. 78 – 89, 2015.
- [8] A. Tri, J. Harjanta, and F. M. Dewanto, “Rancang Bangun Sistem Informasi Penghitungan Pajak Penghasilan (PPh) Pegawai UPGRIS Berbasis Web,” pp. 170 – 176.
- [9] Tang, B., Kay, S. dan Be, H. 2016. Toward Optimal Feature Selection in Naïve Bayes for Text Categorization. *IEEE Transactions on Knowledge and Data Engineering* Vol 28 No 9 Tahun 2016
- [10] Vijayarani, S. dan Dhayanand, S. 2015. *International Journal of Science, Engineering and Technology Research (IJESTR)* Vol 4 No 4 Tahun 2015
- [11] Xu, S. 2016. Bayesian Naïve Bayes Classifiers to Text Classification. *Journal of Information Science* Vol 1 No 12 Tahun 201