

Klasifikasi Penyakit Menular Seksual Menggunakan Naïve Bayes

Gusti Eka Yuliastuti¹, Citra Nurina Prabiantissa², Agung Mustika Rizki³

¹²Jurusan Teknik Informatika, Fakultas Teknik Elektro dan Teknologi Informasi, Institut Teknologi Adhi Tama Surabaya

²Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur

Email: 1gustiekay@itats.ac.id, 2citraturina@itats.ac.id, 3agung.mustika.if@upnjatim.ac.id

Abstract. *The number of sufferers of Sexually Transmitted Diseases (STD) in Indonesia is starting to increase. One example of the case is in the city of Malang, in 2014 as many as 466 people suffer from HIV and 14 people suffer from syphilis. According to the Malang city health report, the average patient is 25 to 49 years old. Some are asymptomatic, where the patient does not feel any symptoms and is not even detected until a medical examination is carried out. In detecting this PMS can use information technology. One way is to build an expert system that applies the Naïve Bayes algorithm to help classify the STD you suffer based on the symptoms you feel. The stages carried out in this research are literature study, data collection, data pre-processing, classification and evaluation. The accuracy results obtained in this study amounted to 76.67%.*

Keywords: *Naïve Bayes, Sexually Transmitted Diseases*

Abstrak. *Angka penderita Penyakit Menular Seksual (PMS) di Indonesia mulai meningkat. Salah satu contoh kasusnya ada di kota Malang, pada tahun 2014 sebanyak 466 orang menderita HIV dan 14 orang menderita Sifilis. Menurut laporan kesehatan kota Malang, rata-rata penderita berusia 25 hingga 49 tahun. Ada pula yang asimtomatik, dimana penderita tidak merasakan gejala apapun bahkan tidak terdeteksi hingga melakukan pemeriksaan medis. Dalam mendeteksi PMS ini dapat memanfaatkan teknologi informasi. Salah satu caranya yakni dengan membangun sistem pakar yang menerapkan algoritma Naïve Bayes untuk dapat membantu mengklasifikasi PMS yang diderita berdasarkan gejala yang dirasakan. Tahapan yang dilakukan dalam penelitian ini yakni studi literatur, pengumpulan data, pre-processing data, klasifikasi dan evaluasi. Adapun hasil akurasi yang didapatkan pada penelitian ini adalah sebesar 76,67%.*

Kata Kunci: *Naïve Bayes, Penyakit Menular Seksual*

1. Pendahuluan

Angka penderita Penyakit Menular Seksual (PMS) di Indonesia mulai meningkat. Penyakit ini sangat erat kaitannya dengan beberapa infeksi lain seperti HIV (Human Immunodeficiency Virus), HPV (Human Papillomavirus), HSV (Herpes Simplex Virus), Hepatitis B, Sifilis, Gonore, Chlamydia dan Trikomoniasis (Aditama, 2011) (World Health Organization, 2016). Salah satu contoh kasusnya ada di kota Malang, pada tahun 2014 sebanyak 466 orang menderita HIV dan 14 orang menderita Sifilis. Menurut laporan kesehatan kota Malang, rata-rata penderita berusia 25 hingga 49 tahun. Ada pula yang asimtomatik, dimana penderita tidak merasakan gejala apapun bahkan tidak terdeteksi hingga melakukan pemeriksaan medis (Dinas Kesehatan Kota Malang, 2015).

Kota Malang kini mulai berkembang dari segi demografis, sosial, pendidikan hingga wisata (Yuliastuti et al., 2017a) (Yuliastuti et al., 2017b). Perkembangan ini juga berbanding lurus dengan resiko penyebaran PMS. Disamping itu, perawatan penyakit ini membutuhkan biaya yang tidak sedikit. Oleh karena itu, penting bagi kita untuk memeriksa diri sendiri lebih awal agar dapat mengurangi beban pengeluaran perawatan.

Seiring dengan perkembangan di era modern ini, untuk mendeteksi PMS ini dapat memanfaatkan teknologi informasi. Salah satunya yakni membangun sistem pakar yang dapat membantu mengklasifikasi PMS yang diderita berdasarkan gejala yang dirasakan.

Pada tahun 2018, terdapat penelitian terkait PMS di Kota Malang (Yuliastuti et al., 2018) dengan menerapkan 3 metode *data mining*. Pada penelitian tersebut, terdapat sebanyak 139 data yang kemudian dibagi menjadi dua yakni untuk kebutuhan *training* data dan *testing* data. Mekanisme *training* menggunakan sebanyak 109 data sedangkan untuk *testing* menggunakan sebanyak 30 data. Selanjutnya hasil dari sistem tersebut akan dibandingkan dengan data sebenarnya dari basis pengetahuan pakar untuk mengetahui akurasi dari penerapan metode *data mining*.

Rizki dkk. (Rizki et al., 2021) melakukan penelitian dengan topik penelitian yang sama seperti penelitian sebelumnya, yakni mengklasifikasi PMS. Namun terdapat perbedaan pada metode yang

diterapkan. Pada penelitian tersebut, penulis menerapkan *Artificial Neural Network*. Kesimpulan dari penelitian tersebut yakni nilai terkecil RMSE didapatkan saat jumlah *hidden nodes* sebanyak 9 buah, nilai *learning rate* sebesar 0,1 dan iterasi yang dilakukan sebanyak 7000 kali. Selain itu Rizki juga menerapkan salah satu metode *data mining* yaitu K-Nearest Neighbour (K-NN) dan menghasilkan akurasi sebesar 90% (Rizki, 2021).

Penelitian terkait klasifikasi penyakit lain juga telah dilakukan, salah satu contohnya adalah penyakit jantung dimana data tersebut didapatkan dari *UCI Machine Learning data set* (Bianto et al., 2020). Penulis menyatakan bahwa dengan menerapkan algoritma Naïve Bayes didapatkan nilai hasil akurasi dengan rata-rata akurasi sebesar 90,61%, rata-rata hasil nilai presisi sebesar 87,44% dan rata-rata nilai recall sebesar 87,95%.

Klasifikasi penyakit Diabetes Mellitus telah dilakukan dengan menerapkan algoritma Naïve Bayes (Ridwan, 2020). Algoritma Naïve Bayes diterapkan oleh penulis dan hasil klasifikasi yang didapatkan yakni sebesar 90,2%. Hasil klasifikasi tersebut menunjukkan bahwa pemodelan algoritma Naïve Bayes terhadap *data set stage Diabetes Mellitus* sudah bagus.

Studi komparasi algoritma telah dilakukan oleh Nurdiana dan Algifari (Nurdiana & Algifari, 2020). Penulis melakukan klasifikasi penyakit Diabetes Mellitus dengan membandingkan penerapan metode ID3 dan Naïve Bayes. Kesimpulan dari penelitian tersebut adalah tingkat akurasi yang dihasilkan oleh penerapan algoritma Naïve Bayes lebih baik dibandingkan dengan ID3.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, penulis akan menerapkan algoritma Naïve Bayes untuk melakukan klasifikasi PMS. Permasalahan PMS ini penting untuk diselesaikan mengingat akan berakibat fatal dan sangat berbahaya jika tidak terdeteksi dan ditangani sedini mungkin.

2. Tinjauan Pustaka

2.1. Penyakit Menular Seksual (PMS)

PMS ini merupakan suatu penyakit yang ditularkan dari satu orang ke orang lain melalui kontak seksual. Akibat dari penyakit ini salah satunya adalah gangguan saluran reproduksi. Jika PMS tidak segera ditangani maka akan berakibat lebih fatal dan menyebabkan penderita sakit berkepanjangan. Beberapa gejala yang dialami oleh penderita PMS antara lain timbul rasa sakit saat buang air kecil, adanya benjolan, terjadi pembengkakan hingga keluar cairan atau nanah dari kemaluan.

2.2. Data Mining

Data mining merupakan suatu ilmu yang mempelajari terkait penggalian data dan menganalisisnya (Yulianti et al., 2018). Adapun yang dapat dianalisis antara lain atribut data, hubungan antar data, pola data, informasi dan pengetahuan berdasarkan data yang ada. *Data mining* bertujuan untuk menemukan informasi dari kumpulan data yang ada. *Data mining* dapat menangani data berskala besar serta dapat memanfaatkan data pengalaman di masa lampau untuk meningkatkan proses model pembelajarannya (Ginantra et al., 2021).

2.3. Naïve Bayes

Naïve Bayes adalah sebuah metode *classifier* yang berdasarkan probabilitas dan Teorema Bayesian, dengan asumsi bahwa setiap variabel bersifat bebas. Dengan kata lain, Naïve Bayes mengasumsikan bahwa keberadaan sebuah atribut tidak ada kaitannya dengan atribut yang lain (Jefi et al., 2021).

Teorema Bayesian merupakan pendekatan statistik dasar untuk pengenalan pola. Teorema Bayesian ini digunakan untuk melakukan perhitungan probabilitas dari suatu hipotesis dari probabilitas *prior* yang diberikan (Shafarindu et al., 2021). Persamaan Teorema Bayesian ditunjukkan sebagai berikut:

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)} \quad (1)$$

Dimana $P(x)$ disebut *prior*, yaitu pengetahuan atau asumsi awal kita. Setelah kita mengobservasi fakta baru (dapat berupa sekumpulan data atau satu *data point*), kita dapat mengubah asumsi kita.

$P(x|y)$ disebut *likelihood function*. *Likelihood function* ini mendeskripsikan peluang data, untuk asumsi atau pengetahuan tentang x yang berubah-ubah (x sebagai parameter yang dapat diatur). Dengan *likelihood function* tersebut, kita mengoreksi pendapat akhir kita yang dapat digunakan untuk mengambil keputusan (*posterior*). Jadi secara umum, Teorema Bayesian ini mengubah *prior* menjadi *posterior* akibat adanya kepercayaan baru (*likelihood*).

Naïve Bayes merupakan algoritma *supervised learning* yang sangat sederhana. Secara formal, persamaan Naïve Bayes untuk klasifikasi adalah sebagai berikut:

$$likelihood(c_i) = P(c_i) \prod_{f=1}^F P(t_f|c_i) \quad (2)$$

Dimana c_i adalah suatu nilai kelas, C adalah kelas (himpunan), t adalah fitur (satu fitur, bukan *feature vector*) dan F adalah banyaknya fitur. Dengan *likelihood*, kita memprediksi kelas berdasarkan probabilitas kemunculan nilai fitur pada kelas tersebut. Pertama kita menghitung *likelihood* suatu *feature vector* diklasifikasikan ke kelas tertentu berdasarkan bagaimana probabilitas korespondensi fitur-fiturnya terhadap kelas tersebut menggunakan persamaan yang sama seperti sebelumnya. Kemudian normalisasi *likelihood* semua kelas untuk mendapatkan probabilitas *class-assignment (softmax)* dengan menggunakan persamaan berikut:

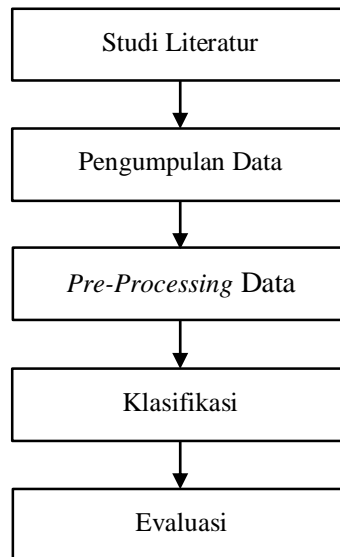
$$P_{assignment}(c_i) = \frac{likelihood(c_i)}{\sum_{c_j \in C} likelihood(c_j)} \quad (3)$$

Setelah itu, kita memilih kelas dengan probabilitas tertinggi, dengan menggunakan persamaan berikut:

$$\hat{c}_i = \arg_{c_i \in C} P_{assignment}(c_i) \quad (4)$$

3. Metode Penelitian

Dalam melaksanakan penelitian diperlukan pedoman berupa tahapan yang terstruktur agar mendapat hasil penelitian sesuai yang diharapkan. Adapun tahapan dalam penelitian ini seperti ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

3.1. Studi Literatur

Tahap pertama dalam penelitian ini yakni melakukan studi literatur. Tahap ini dilakukan dengan memahami penelitian-penelitian terdahulu untuk mendapatkan teori yang mendukung dan informasi yang dapat digunakan dalam penelitian ini. Adapun literatur yang dipahami berkaitan dengan Penyakit Menular Seksual (PMS) dan algoritma Naïve Bayes.

3.2. Pengumpulan Data

Tahap kedua yang perlu dilakukan adalah mengumpulkan data. Data yang akan digunakan dalam penelitian ini didapat dari salah satu puskesmas di Kota Malang sejumlah 139 contoh data. Contoh data tersebut dibagi menjadi dua, yakni 109 data sebagai data latih dan 30 data sebagai data uji. Terdapat beberapa jenis PMS antara lain (Yuliastuti et al., 2018): *Gonorrhoea*, *Syphilis*, *Herpes Genitalis*, *Chlamydia*, *Trichomoniasis Vaginalis*, *Kutil Kelamin*, *Chankroid*, *Granuloma Inguinale*, *Limfogranulama Venereum*, *Servictis*, *Kandidiasis Vulvavaginalis*, *Vaginiosis Bacterial*, *Moluscum Kontagiosum*, *Proctitis*, *Konjungtivitis Neonatorum* dan Radang Panggul.

3.3. Pre-Processing Data

Tahap ketiga yang perlu dilakukan yakni *pre-processing* data, diaman ini merupakan pengolahan data awal. Dalam tahap ini akan menghilangkan data-data yang tidak sesuai, menghapus data ganda serta menghapus bagian data yang kosong.

3.4. Klasifikasi

Setelah mengolah data awal, tahap selanjutnya adalah melakukan klasifikasi dengan menggunakan algoritma Naïve Bayes. Dengan menerapkan algoritma Naïve Bayes, nantinya perlu melakukan beberapa proses perhitungan untuk mendapatkan hasil akhir.

3.5. Evaluasi

Tahap terakhir dalam penelitian ini yakni melakukan evaluasi untuk mengetahui akurasi dari penerapan algoritma Naïve Bayes.

4. Hasil dan Pembahasan

Pada penelitian ini, pengujian dilakukan terhadap 30 *data testing* dimana sebelumnya telah dilakukan pembelajaran oleh 139 *data training*. Hasil dari dari pengujian tersebut kemudian dibandingkan dengan data sebenarnya berdasarkan basis pengetahuan dari pakar, guna mengetahui akurasi dari penerapan algoritma Naïve Bayes ini. Hasil pengujian seperti ditunjukkan pada Tabel 1.

Tabel 1. Hasil Pengujian

No	Basis Pengetahuan Pakar	Hasil Sistem
1	Gonorrhoea	Trichomoniasis Vaginalis
2	Moluskum Kontagiosum	Moluskum Kontagiosum
3	Moluskum Kontagiosum	Moluskum Kontagiosum
4	Proctitis	Proctitis
5	Konjungtivitis neonatorum	Konjungtivitis Neonatorum
6	Bacterial Vaginosis	Trichomoniasis Vaginalis
7	Limfogranulama Venereum	Limfogranulama Venereum
8	Granuloma Inguinale	Granuloma Inguinale
9	Syphilis	Syphilis
10	Kandidiasis	Kandidiasis
11	Trichomoniasis Vaginalis	Vaginal Candidiasis
12	Trichomoniasis Vaginalis	Vaginal Candidiasis
13	Servictis	Servictis

No	Basis Pengetahuan Pakar	Hasil Sistem
14	Syphilis	Syphilis
15	Gonorrhoea	Trichomoniasis Vaginalis
16	Konjungtivitis Neonatorum	Konjungtivitis Neonatorum
17	Radang Panggul	Radang Panggul
18	Pembengkakan Skrotum	Pembengkakan Skrotum
19	Bacterial Vaginosis	Trichomoniasis Vaginalis
20	Servicitis	Servicitis
21	Chlamydia	Chlamydia
22	Herpes Genitalis	Herpes Genitalis
23	Granuloma Inguinale	Granuloma Inguinale
24	Kutil Kelamin	Kutil Kelamin
25	Bacterial Vaginosis	Trichomoniasis Vaginalis
26	Chancroid	Chancroid
27	Herpes Genitalis	Herpes Genitalis
28	Proctitis	Proctitis
29	Chlamydia	Chlamydia
30	Chlamydia	Chlamydia

Berdasarkan tabel hasil pengujian, ditemukan sebanyak 7 *data testing* yang tidak sesuai dengan basis pengetahuan dari pakar, sehingga akurasi dari penelitian ini sebagai berikut:

$$Akurasi = \frac{23}{30} \times 100\% = 76,67\% \quad (5)$$

Ketidakesuaian tujuh data tersebut disebabkan oleh ada lebih dari satu gejala yang mirip sehingga mengakibatkan terjadinya kesalahan klasifikasi satu penyakit dengan penyakit lainnya.

5. Kesimpulan

Berdasarkan hasil pengujian tersebut, maka didapatkan kesimpulan bahwa penerapan algoritma Naïve Bayes pada penelitian ini cukup optimal untuk menyelesaikan permasalahan klasifikasi dengan akurasi sebesar 76,67%. Untuk penelitian selanjutnya, penulis perlu meningkatkan akurasi tersebut dengan mengoptimalkan penerapan algoritma Naïve Bayes. Optimasi metode Naïve Bayes juga dapat dilakukan dengan menggabungkan algoritma optimasi yang lain. Disamping itu, penulis juga perlu menambahkan *data training* agar akurasi hasilnya lebih baik.

Referensi

- Aditama, T. Y. (2011). *Pedoman Nasional Penanganan Infeksi Menular Seksual*. Kementerian Kesehatan Republik Indonesia.
- Bianto, M. A., Kusri, K., & Sudarmawan, S. (2020). Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. *Creative Information Technology Journal*, 6(1), 75. <https://doi.org/10.24076/citec.2019v6i1.231>
- Dinas Kesehatan Kota Malang. (2015). *Profil Kesehatan Kota Malang Tahun 2014*.
- Ginantra, N. L. W. S. R., Arifah, F. N., Wijaya, A. H., Septarini, R. S., Ahmad, N., Ardiana, D. P. Y. A., Effendy, F., Iskandar, A., Hazriani, H., Sari, I. Y., Gustiana, Z., Prianto, C., Gustian, D., & Negara, E. S. (2021). *Data Mining dan Penerapan Algoritma* (R. Watianthos & J. Simarmata (eds.)). Yayasan Kita Menulis.
- Jefi, J., Hendri, H., Afni, N., Salim, A., & Maulana, Y. I. (2021). Penerapan Algoritma Naive Bayes dalam Memprediksi Penyakit Lambung. *Journal of Information System, Informatics and Computing*, 5(2). <https://doi.org/10.52362/jisicom.v5i2.659>
- Nurdiana, N., & Algifari, A. (2020). Studi Komparasi Algoritma ID3 dan Algoritma Naive Bayes untuk

- Klasifikasi Penyakit Diabetes Mellitus. *INFOTECH Journal*, 6(2), 18–23.
<https://ejournal.unma.ac.id/index.php/infotech/article/view/816>
- Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 15–21.
<https://doi.org/10.47970/siskom-kb.v4i1.169>
- Rizki, A. M. (2021). Sistem Pakar Diagnosa Penyakit Menular Seksual Menggunakan K-NN. *SCAN - Jurnal Teknologi Informasi Dan Komunikasi*, 16(2), 2–5.
<https://doi.org/10.33005/scan.v16i2.2691>
- Rizki, A. M., Maulana, H., Kartika, D. S. Y., & Yuliastuti, G. E. (2021). Classification Of Sexually Transmitted Infectious Diseases Using Artificial Neural Networks. *Jurnal Mantik*, 5(36), 1759–1765.
- Shafarindu, A. I., Ernawati, L., & Zaidiah, A. (2021). Penerapan Algoritma Naive Bayes untuk Klasifikasi Tingkat Kebugaran Jasmani Berdasarkan Hasil Pengukuran pada Pegawai. *Seminar Nasional Mahasiswa Ilmu Komputer Dan Aplikasinya (SENAMIKA)*, September, 278–287.
- World Health Organization. (2016). *Media Centre-Sexually Transmitted Infections(STIs)*.
<http://www.who.int/mediacentre/factsheets/fs110/en/>
- Yuliastuti, G. E., Alfiyatin, A. N., Rizki, A. M., Hamdianah, A., Taufiq, H., & Mahmudy, W. F. (2018). Performance Analysis of Data Mining Methods for Sexually Transmitted Disease Classification. *International Journal of Electrical and Computer Engineering*, 8(5), 3933–3939.
<https://doi.org/10.11591/ijece.v8i5.pp3933-3939>
- Yuliastuti, G. E., Mahmudy, W. F., & Rizki, A. M. (2017a). Implementation of Genetic Algorithm to Solving Travelling Salesman Problem with Time Window (TSP-TW) for Scheduling Tourist Destinations in Malang City. *Journal of Information Technology and Computer Science*, 2(1), 1–10.
- Yuliastuti, G. E., Mahmudy, W. F., & Rizki, A. M. (2017b). Penanganan Fuzzy Time Window pada Travelling Salesman Problem (TSP) dengan Penerapan Algoritma Genetika. *MATICS Jurnal Ilmu Komputer Dan Teknologi Informasi*, 9(1), 38–43.