

# PREDIKSI KLASIFIKASI PERAWATAN PADA DATASET KANKER PAYUDARA COIMBRA DENGAN METODE NAIVE BAYES

Ferawaty<sup>1</sup>, Wenripin Chandra<sup>2</sup>, Kelvin Ivanka<sup>3</sup>

<sup>1</sup> Fakultas Ilmu Komputer, Universitas Pelita Harapan Kampus Medan  
E-mail: ferawaty.fik@UPH365.onmicrosoft.com

<sup>2</sup> Fakultas Ilmu Komputer, Universitas Pelita Harapan Kampus Medan  
E-mail: wenripin@lecturer.uph.edu

<sup>2</sup> Fakultas Ilmu Komputer, Universitas Pelita Harapan Kampus Medan  
E-mail: ki70033@student.uph.edu

---

*Abstract – Breast cancer is a dreaded disease and a major cause of death. In this study, the Naïve Bayes method is used to predict the category of breast cancer treatment for the Breast Cancer Coimbra Dataset. Test results involving nine variables in the dataset resulted in 44.8% of the "Healthy Controls" category and 55.2% of the "Patient" category.*

**Keywords:** Breast Cancer, Naive Bayes, Coimbra, Classification

*Abstrak – Kanker payudara merupakan penyakit yang ditakuti dan menjadi penyebab kematian yang besar. Dalam penelitian ini, metode Naïve Bayes digunakan untuk memprediksi kategori perawatan kanker payudara terhadap Breast Cancer Coimbra Dataset. Hasil pengujian yang melibatkan sembilan variabel dalam dataset tersebut menghasilkan 44,8% berkategori "Healthy Controls" dan 55,2% berkategori "Patient".*

**Kata Kunci:** Kanker Payudara, Naive Bayes, Coimbra, Klasifikasi

---

## PENDAHULUAN

Kanker payudara merupakan tumor ganas yang menyerang jaringan payudara yang berasal dari kelenjar, saluran kelenjar dan jaringan penunjang payudara. Kanker payudara terjadi karena adanya kerusakan gen yang mengatur pertumbuhan dan diferensiasi sehingga sel-sel tersebut tumbuh dan bertambah tanpa dapat dikendalikan [1]. Kanker ini diderita oleh sekitar 2,1 juta wanita setiap tahun dan pada tahun 2018 diperkirakan menjadi penyebab kematian pada 627.000 wanita penderitanya, kira-

kira 15% dari semua kematian pada wanita akibat kanker [2]. Kondisi seperti ini membuat diagnosis awal penyakit menjadi penting untuk dilakukan.

Pada penelitian ini diagnosis dilakukan dengan mengaplikasikan metode Naïve Bayes terhadap data tes darah yang dilakukan di Coimbra, Portugal untuk membuat prediksi terjadinya kanker payudara.

## METODE PENELITIAN

Metode Naïve Bayes merupakan metode klasifikasi menggunakan probabilitas

dan statistik untuk memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Metode ini mengasumsikan independensi masing-masing variabel atau indikator yang digunakan dalam pengolahan [3]. Metode ini memiliki akurasi yang lebih baik dibanding metode pohon keputusan dan *neural network* dalam pengklasifikasian halaman-halaman *web* [4].

Tahapan proses metode Naïve Bayes adalah:

- Menghitung jumlah kelas/label
- Menghitung jumlah kasus per kelas
- Kalikan semua variabel kelas
- Bandingkan hasil per kelas.

Persamaan Naïve Bayes yaitu:

$$P(C|X) = P(x|c)P(c)P(x) \quad (1)$$

dimana:

*x* : Data dengan kelas yang belum diketahui

*c* : Hipotesis data merupakan suatu kelas spesifik

*P(c/x)* : Probabilitas hipotesis berdasar kondisi (*posterior probability*)

*P(c)* : Probabilitas hipotesis (*prior probability*)

*P(x/c)* : Probabilitas berdasarkan kondisi pada hipotesis

*P(x)* : Probabilitas *c*

Data yang digunakan dalam penelitian ini adalah *dataset* Breast Cancer Coimbra [5]. *Dataset* ini memiliki sembilan variabel/indikator yaitu:

a. Usia (*age*)

Bagi yang berusia di bawah 25 tahun, usia 32 tahun adalah batasan orang dikatakan muda. Sementara usia yang tergolong tua adalah yang berusia di atas 54 tahun [6].

b. *Body Mass Index* (BMI)

BMI di bawah 18,5 tergolong *underweight*, di antara 18,5 hingga 24,9 termasuk sehat sedangkan BMI di antara 25 hingga 29,9 termasuk *overweight* [7].

c. Glukosa (*glucose*)

Tingkat normal glukosa dalam darah dalam kondisi puasa untuk bukan penderita penyakit kencing manis berkisar 3,9 dan 7,1 mmol/L (70 hingga 130 mg/dL) [8].

d. Insulin

Kadar insulin puasa adalah antara 3–8 uIU/mL (18–48 pmol/L) [9].

e. *HOMA*

Nilai normal *HOMA-IR* yang sehat berkisar antara 0,5-1,4. Nilai yang lebih kecil dari 1,0 menunjukkan keoptimalan pada tingkat sensitivitas insulin [10].

f. *Leptin*

Kadar *leptin* yang normal bagi perempuan adalah antara 18,7 ng/mL hingga 28,4 ng/mL [11].

g. *Adiponectin*

Kadar *adiponectin* selama masa kehamilan berkisar antara 2.7 µg/mL hingga 25.0 µg/mL, sedangkan tidak dalam masa kehamilan kadarnya antara 3.5 µg/mL dan 22.4 µg/mL [12].

h. *Resistin*

Kadar *resistin* normal berkisar antara 4-12 ng/mL [13].

i. MCP-1 (*Monocyte Chemoattractant Protein-1*)

Kadar MCP-1 wanita sehat adalah 67–425 pg/mL, sedangkan untuk penderita kanker payudara adalah 57-692 pg/mL [14].

Gejala-gejala ini dikategorikan untuk menentukan apakah peserta tersebut hanya menjalankan kontrol kesehatan (*Healthy Controls*, diberi label 1) atau harus menjalani perawatan sebagai pasien kanker (*Patient*, diberi label 2).

Sampel data penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Sampel Data Penelitian

Usia	Insulin	<i>Leptin</i>	<i>Resistin</i>	Klasifikasi
48	2,707	8,8071	7,99585	1
25	4,09	20,45	5,14	1
61	18,077	30,7729	13,68392	1
59	3,188	17,022	31,6904	2

72	2,82	24,96	3,27	2
34	3,469	14,57	6,92	1
29	5,81	45,6196	24,6033	1
45	6,76	39,9802	8,70448	1
60	30,13	37,843	11,50005	2
45	4,713	23,8479	15,55625	2

Hasil kategorisasi tersebut terdapat dalam Tabel 2.

Tabel 2. Data setelah Kategorisasi

Usia	Insulin	Leptin	Resistin	Klasifikasi
Dewasa	Low	Low	Normal	Healthy Controls
Muda	Normal	Normal	Normal	Healthy Controls
Tua	High	High	High	Healthy Controls
Tua	Normal	Low	High	Patient
Tua	Low	Normal	Low	Patient
Dewasa	Normal	Low	Normal	Healthy Controls
Muda	Normal	High	High	Healthy Controls
Dewasa	Normal	High	Normal	Healthy Controls
Tua	High	High	Normal	Patient
Dewasa	Normal	Normal	High	Patient

Data yang telah dikategorikan akan diuji prediksi dengan tabel data uji seperti pada Tabel 3.

Tabel 3. Data Uji

Usia	Insulin	Leptin	Resistin	Klasifikasi
Tua	High	Low	High	??

Langkah-langkah pengujian:

**P(C)**

$$P(\text{Klasifikasi} = \text{"Healthy Controls"}) = 6/10$$

$$P(\text{Klasifikasi} = \text{"Patient"}) = 4/10$$

**P(X|C)**

- $P(\text{Usia} : \text{"Tua"} \mid \text{Klasifikasi} : \text{"Healthy Controls"}) = 1/6$

$$P(\text{Usia} : \text{"Tua"} \mid \text{Klasifikasi} : \text{"Patient"}) = 3/4$$

- $P(\text{Insulin} : \text{"High"} \mid \text{Klasifikasi} : \text{"Healthy Controls"}) = 1/6$

$$P(\text{Insulin} : \text{"Normal"} \mid \text{Klasifikasi} : \text{"Patient"}) = 1/4$$

- $P(\text{Leptin} : \text{"Low"} \mid \text{Klasifikasi} : \text{"Healthy Controls"}) = 2/6$

$$P(\text{Leptin} : \text{"Low"} \mid \text{Klasifikasi} : \text{"Patient"}) = 1/4$$

- $P(\text{Resistin} : \text{"High"} \mid \text{Klasifikasi} : \text{"Healthy Controls"}) = 2/6$

$$P(\text{Resistin} : \text{"High"} \mid \text{Klasifikasi} : \text{"Patient"}) = 2/4$$

$$P(X \mid \text{Klasifikasi} : \text{"Healthy Controls"}) = 1/6 \times 1/6 \times 2/6 \times 2/6$$

$$= 0,0038064198 \approx 0,0038$$

$$P(X \mid \text{Klasifikasi} : \text{"Patient"})$$

$$= 3/4 \times 1/4 \times 1/4 \times 2/4$$

$$= 0,0234375 \approx 0,0234$$

**P(X|C) \* P(C)**

$$P(X \mid \text{Klasifikasi} : \text{"Healthy Controls"}) \times P(\text{Klasifikasi} : \text{"Healthy Controls"})$$

$$= 0,0038 \times 6/10$$

$$= 0,00228$$

$$P(X \mid \text{Klasifikasi} : \text{"Patient"}) \times$$

$$P(\text{Klasifikasi} : \text{"Patient"})$$

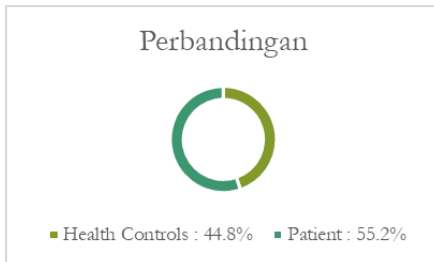
$$= 0,0234 \times 4/10$$

$$= 0,00936$$

Karena **0,00936 > 0,00228**, maka peserta pada data uji tersebut termasuk dalam klasifikasi **"Patient"**

### HASIL DAN PEMBAHASAN

Dalam penelitian ini, setiap data yang terdapat pada *dataset* yang digunakan diuji dengan metode Naïve Bayes dan menghasilkan data yang diklasifikasikan sebagai *"Health Control"* atau *"Patient"*. Dari 116 data pada *dataset* yang diuji, diperoleh hasil sebanyak 52 data (44,8%) berklasifikasi *"Healthy Controls"* dan sisanya 64 data (55,2%) diklasifikasikan sebagai *"Patient"*.



Gambar 1. Perbandingan Jumlah Kategori yang Dihasilkan dari Pengujian Metode Penggunaan metode Naïve Bayes memberikan jumlah per kategori yang sama dengan hasil yang dituliskan pada *dataset* yang digunakan.

### KESIMPULAN

Kesimpulan yang didapat dari hasil pengujian adalah:

1. Metode Naïve Bayes digunakan untuk mengklasifikasikan data penderita kanker payudara.
2. Metode Naïve Bayes memberikan jumlah per kategori yang sama saat digunakan untuk mengklasifikasikan data penderita kanker payudara.

### DAFTAR PUSTAKA

- [1] L. Mardiana, Kanker Pada Wanita Pencegahan dan Pengobatan dengan Tanaman Obat, Jakarta: Penebar Swadaya, 2004.
- [2] "Breast Cancer Early Diagnosis and Screening," World Health Organization, [Online]. Available: <https://www.who.int/cancer/prevention/diagnosis-screening/breast-cancer/en/>.
- [3] "Algoritma Naive Bayes," 15 7 2017. [Online]. Available: <https://informatikalogi.com/algoritma-naive-bayes/>.
- [4] D. Xhemali, C. J. Hinde and R. G. Stone, "Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages," *International Journal of Computer Science*, vol. 4, no. 1, pp. 16-23, 2009.
- [5] M. Patrício, J. Pereira, J. Crisóstomo, P. Matafome, M. Gomes, R. Seça and F. Caramelo, "Breast Cancer Coimbra Data Set," Januari 2018. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Coimbra>.
- [6] "Batasan Usia Disebut Muda dan Tua," 13 1 2012. [Online]. Available: <https://www.beritasatu.com/beritasatu-administrator/archive/25895/batasan-usia-disebut-muda-dan-tua>.
- [7] "What is the body mass index (BMI)?," [Online]. Available: <https://www.nhs.uk/common-health-questions/lifestyle/what-is-the-body-mass-index-bmi/>.
- [8] "Blood sugar level," [Online]. Available: [https://en.wikipedia.org/wiki/Blood\\_sugar\\_level](https://en.wikipedia.org/wiki/Blood_sugar_level).
- [9] "Insulin Resistance/T2 Diabetes: Map Your Test Results," [Online]. Available: <https://www.thebloodcode.com/insulin-resistancet2-diabetes-map-test-results/>.
- [10] "What is the normal HOMA-IR and HOMA-β for the healthy human?," 24 9 2014. [Online]. Available: [https://www.researchgate.net/post/What\\_is\\_the\\_normal\\_HOMA-IR\\_and\\_HOMA-b\\_for\\_the\\_healthy\\_human](https://www.researchgate.net/post/What_is_the_normal_HOMA-IR_and_HOMA-b_for_the_healthy_human).
- [11] T. Gijón-Conde, A. Graciani, P. Guallar-Castillón, M. T. Aguilera, F. Rodríguez-Artalejo and J. R. Banegas, "Leptin Reference Values and Cutoffs for Identifying Cardiometabolic Abnormalities in the Spanish Population," *Revista Española de Cardiología*, vol. 68, no. 8, pp. 672-679, August 2015.
- [12] J. K. Nien, S. Mazaki-Tovi, R. Romero, O. Erez, J. P. Kusanovic, F. Gotsch, B. L. Pineles, R. Gomez, S. Edwin, M. Mazor, J. Espinoza, B.

- H. Yoon and S. S. Hassan, "Plasma Adiponectin Concentrations in Non Pregnant, Normal Pregnancy and Overweight Pregnant Women," *Journal of Perinatal Medicine*, vol. 35, no. 6, 4 6 2008.
- [13] D. A. De Luis, M. G. Sagrado, R. Conde, R. Aller and O. Izaola, "Resistin Levels and Inflammatory Markers in Patients with Morbid Obesity," *Nutricion Hospitalaria*, vol. 25, no. 4, pp. 630-634, 2010.
- [14] A. Lebrecht, C. Grimm, T. Lantzsch, E. Ludwig, L. Hefler, E. Ulbrich and H. Koelbl, "Monocyte Chemoattractant Protein-1 Serum Levels in Patients with Breast Cancer," *Tumor Biology*, vol. 25, no. 1-2, p. 14-17, 6 2004.