

ANALISIS KLASIFIKASI RISIKO TERHADAP PENDERITA PROSTAT MENGGUNAKAN METODE NAIVE BAYES

Indra Saputra¹, Bakhtiyar Hadi P²

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember
osingputra@gmail.com, bahtiyar.task@gmail.com

ABSTRAK

Kelenjar prostat merupakan organ tubuh pada laki-laki yang berbentuk seperti kacang kenari, kelenjar prostat terletak di dasar kandung kemih dan mengelilingi *uretra posterior*, salah satu gangguan pada prostat adalah terjadinya pembesaran yang lazimnya terjadi pada pria di atas 50 tahun. Di Indonesia BPH menjadi penyakit urutan ke dua dengan jumlah penderita terbanyak setelah penyakit batu saluran kemih, dan secara umum diperkirakan hampir 50% pria Indonesia menderita BPH, jika dilihat dari 200 juta lebih rakyat Indonesia maka dapat di perkirakan sekitar 2,5 juta pria yang berumur lebih dari 60 tahun menderita BPH (Purnomo, 2008). Penyakit ini perlu diwaspadai karena bila tidak segera ditangani dapat mengganggu saluran kemih, efek jangka panjang yang timbul adalah retensi urine akut, *refluks* kandung kemih, *hidroureter*, dan *urinari tract infection*. Dari penelitian ini mendapatkan hasil hitung dari data uji terhadap confusion matrix menunjukkan bahwa tingkat akurasi 70%, presisi 67% dan recall 40%.

1. PENDAHULUAN

2. TINJAUAN PUSTAKA

2.1. Prostat

Benigna Prostate Hiperplasia (BPH) merupakan perbesaran kelenjar prostat, memanjang ke atas kedalam kandung kemih dan menyumbat aliran urin dengan menutupi orifisium uretra akibatnya terjadi dilatasi ureter (hidroureter) dan ginjal (hidronefrosis) secara bertahap (Smeltzer dan Bare, 2002).

2.2. Sistem kardiovaskuler

Sistem kardiovaskuler merupakan organ sirkulasi darah yang terdiri dari jantung, komponen darah dan pembuluh darah. Fungsi sirkulasi adalah untuk memenuhi kebutuhan jaringan tubuh, untuk mentranspor zat makanan ke seluruh tubuh, untuk mentranspor produk-produk yang tidak berguna, untuk mentranspor oksigen ke seluruh tubuh, untuk menghantarkan hormon dari satu bagian tubuh ke bagian tubuh yang lain, dan secara umum untuk memelihara lingkungan yang sesuai di dalam seluruh cairan jaringan tubuh agar sel bisa bertahan hidup dan berfungsi secara optimal.

2.3. Elektrokardiogram

Elektrokardiogram (EKG) merupakan suatu sinyal yang terbentuk sebagai hasil dari aktivitas listrik jantung. EKG diambil dengan memasang elektroda pada titik tertentu tubuh pasien. Sinyal EKG mempunyai bentuk spesifik sehingga dapat dijadikan sebagai acuan untuk menentukan kondisi kesehatan jantung oleh ahli jantung. Sinyal EKG direkam menggunakan perangkat elektrokardiograf, (Abedin – Conner, Belajar EKG Tanpa Guru, 1991).

2.4. Kadar Hemoglobin

Kadar hemoglobin adalah ukuran pigmen respiratorik dalam butiran-butiran darah merah. Jumlah hemoglobin dalam darah normal adalah kira-kira 15gr setiap 100 ml darah dan jumlah ini disebut “100 persen” (Evelyn, 2008). Hemoglobin adalah protein yang kaya akan zat besi. Mempunyai daya gabung terhadap oksigen dan dengan oksigen itu membentuk oxihemoglobin di dalam sel darah merah. Dengan fungsi ini maka oksigen dibawa dari paru-paru ke jaringan tubuh (Evelyn, 2008).

2.5. Data Mining

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam database. Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar. (Turban, dkk. 2005)

2.6. Naive 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. Naive Bayes juga didefinisikan sebagai pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan

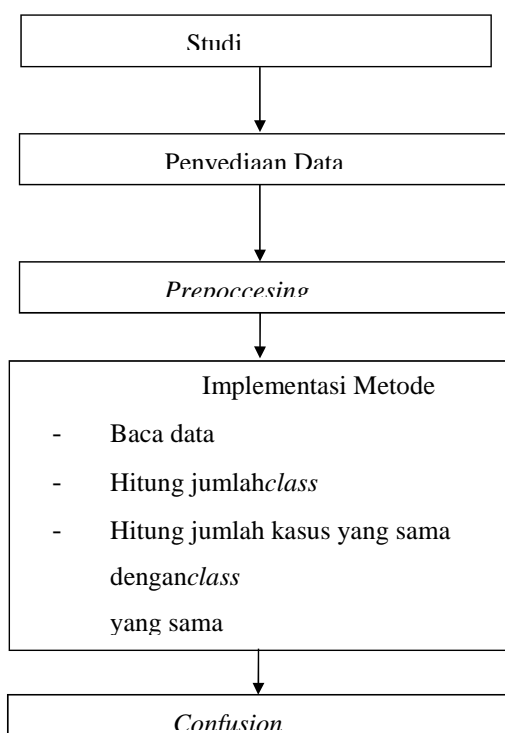
pengalaman di masa sebelumnya (Saleh, 2015).

2.7. Confusion Matrix

Evaluasi hasil klasifikasi dilakukan dengan metode confusion matrix. Confusion matrix adalah tool yang digunakan sebagai evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas sebenarnya atau dengan kata lain berisi informasi nilai sebenarnya dan prediksi pada klasifikasi (Gorunescu, 2011: 319).

3. METODOLOGI PENELITIAN

3.1. Tahapan



3.2. Penyediaan Dataset

Penelitian ini menggunakan dataset yang diambil dari situs internet pada laman web <ftp://ftp.uni-bayreuth.de/pub/math/statlib/s/Harrell/data/xls/prostate.xls> yang berisi beberapa atribut diantaranya; tingkatan penyakit (stadium), konsumsi obat per hari, usia penyakit (bulan), usia pasien, aktivitas per hari, riwayat cardiovascular, electrocardiogram, serum hemoglobin, serum PAP (Prostatic Acid Phosphatase), bone metastases.

3.3. Dsd

No.	Atribut	Data Awal	Preprocessing
1	Tingkatan penyakit	Angka 3-4	Tidak ada Preprocessing

2	Konsumsi obat	0,2 mgestrogen 1,0 mgestrogen 5,0mgestrogen Placebo	0,2 1,0 5,0 Obat lain
3	usia penyakit (bulan)	Angka 1-76	Fase pertama (1-3 tahun), Fase kedua (3-5 tahun) dan Fase ketiga (>3 tahun)
4.	Usia	Angka 48-89	Lansia (46-65 tahun) Manula >65 tahun
5	Serum Hemoglobin	Angka 5-21	Dibawah Normal Normal Hemoglobin 14-18 gr/dl Diatas normal >18
6	Serum Prostatic Acid Phosphatase	Angka 0,09-99	Normal <=2,1 nM Diatas normal >2,1 nM

4. IMPLEMENTASI DAN PENGUJIAN

4.1. Persiapan data

TINGKATAN PENYAKIT	KONSUMSI OBAT	USIA PASIEN	AKTIVITAS	RIWAYAT CARDIOVASCULAR	ELEKTROKARDIOGRAM	SERUM HEMOGLOBIN	SERUM PAP	STATUS
1	0,2	3	normal	tidak ada	tidak ada	dibawah normal	normal	hidup
2	1,0	4	normal	tidak ada	tidak ada	dibawah normal	normal	hidup
3	5,0	5	normal	tidak ada	tidak ada	dibawah normal	normal	hidup
4	Placebo	6	normal	tidak ada	tidak ada	dibawah normal	normal	hidup

						cond ucti on def			
3	3	0.	lan sia	norm al	tid ak	hear t strai n	nor mal	nor mal	menin ggal
4	3	5	ma nul a	norm al	ya	beni gn	diba wah nor mal	nor mal	menin ggal
5	3	0.	ma nul a	dikas ur	ya	nor mal	diat as nor mal	nor mal	menin ggal
6	3	ob at lai n	ma nul a	norm al	tid ak	nor mal	diba wah nor mal	nor mal	hidup
7	3	0.	ma nul a	norm al	tid ak	beni gn	nor mal	nor mal	menin ggal
8	3	ob at lai n	ma nul a	norm al	tid ak	hear t strai n	diba wah nor mal	nor mal	menin ggal
9	3	ob at lai n	ma nul a	norm al	ya	nor mal	diba wah nor mal	nor mal	hidup
1 0	3	1	lan sia	norm al	tid ak	old mi	nor mal	nor mal	hidup

TN	30
FP	4
FN	12
Total	54

Akurasi	70%
Presisi	67%
Recall	40%

5. PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian, Analisis Klasifikasi Risiko Terhadap Penderita Prostat Menggunakan Metode *Naive Bayes* dengan data 506 record, data latih 452, data uji 54 record, diperoleh kesimpulan sebagai berikut :

1. Klasifikasi penderita prostat menggunakan naive bayes didukung dengan sistem yang dibangun menggunakan PHP data database ini memiliki keunggulan lebih cepat dibandingkan dengan klasifikasi manual atau penghitungan manual pada excel misalnya.
2. Hasil hitung dari data uji terhadap confusion matrix menunjukkan bahwa tingkat akurasi 70%, presisi 67% dan recall 40%.

5.2. Saran

Berikut ini adalah saran yang dapat membantu penelitian agar lebih baik dan berkembang, antara lain :

1. Preprocessing pada atribut masih dilakukan secara manual, sehingga perlu adanya pengembangan terhadap proses ini agar dapat mengefisiensi waktu.
2. Pengembang dapat membandingkan dengan metode lain, untuk mencari serta menentukan metode yang paling tepat.
3. Pada sistem ini pasien baru tidak terekam, kepada pengembang data pasien bisa terekam pada database.

4.2. Penghitungan naive bayes

Atribut	Jumlah Kasus	Hidup	Meninggal	C1(Hidup)	C2(Meninggal)
Total	452	137	315	0.30309734513274	0.69690265486726
tingkatan_penyakit					
3	269	92	179	0.67153084671533	0.5825396825397
4	181	46	137	0.33576642335766	0.43492063492063
konsumsi_obat					
0.2	111	30	83	0.21897810218978	0.26349206349206
1.0	113	50	65	0.3649630364964	0.20634920634921
5.0	110	30	82	0.21897810218978	0.26031746031746
obat_lain	116	30	88	0.21897810218978	0.27936507936508
usia_pasien					
manula	370	107	265	0.78102189781022	0.84726984726984
lanusia	79	30	51	0.21897810218978	0.16190476190476
aktivitas					
normal	405	131	276	0.95620437956204	0.87619047619048
dikatur	45	7	40	0.051094890510949	0.12098472698473
riwayat_cardiovascular					
ya	187	38	151	0.27737226277372	0.47936507936508
tidak	263	100	165	0.72992700729927	0.52380952380952
elektrokardiogram					
benign	23	9	16	0.06589340658934	0.050793650793651
heart_strain	142	40	104	0.29197802919781	0.33015873015873
heart_strain	142	40	104	0.29197802919781	0.33015873015873
old_mi	65	20	47	0.14598540145985	0.14820634920635

4.3. Confusion matrix

Kriteria	Jumlah
TP	8

Daftar pustaka

- D. Kandou Manado periode tahun 2013–2015. 2Bagian Ilmu Bedah Fakultas Kedokteran Universitas Sam Ratulangi Manado.
- Gyan Aryadi. (2014). Klasifikasi Keganasan Kanker Prostat Berdasarkan Pada Analisis Citra Biomedis. Jurusan Matematika Universitas Padjadjaran, Bandung.
- Hanaa Ismail Elshazly. (2013). Ensemble-based classifiers for prostate cancer diagnosis. Faculty of Computers and Information, Cairo University, Cairo - Egypt.
- Mitchell H. Gail, MD, PhD. (1995). Dave Byafs Contribution to Epidemiology. E~ostatistics Branch, National Cancer Institute, Rockville, Maryland.

SW Lee. (2017). A Sol-gel Integrated Dual-readout Microarray Platform for Quantification and Identification of Prostate-specific Antigen. Department of Chemistry, University of Tokyo, Tokyo, Japan. Valdo R. Solang dkk. (2016). Profil penderitanya kanker prostat di RSUP Prof. Dr. R. Ying Liu. (2017). A Classification Model for the Prostate Cancer Based on Deep Learning. School of Medical Instrument and Food Engineering University of Shanghai for Science and Technology Shanghai, China.