

SISTEM KLASIFIKASI PENYEBARAN PENYAKIT MATA DI JAWA BARAT DENGAN ALGORITMA ITERATIVE DICHOTOMISER 3 DAN NAÏVE BAYES CLASSIFIER

Wina Witanti^{1*}, Dian Nursantika², Zillan Taufiq Budiman³

^{1,2,3}Program Studi Informatika, Fakultas MIPA, Universitas Jenderal Achmad Yani
Jl. Terusan Jenderal Sudirman, Cimahi, Jawa Barat, 40533

*Email: wina.witanti@lecture.unjani.ac.id

Abstrak

Sistem klasifikasi penyebaran penyakit mata merupakan suatu proses untuk mengetahui potensi penyakit mata di setiap daerah di Jawa Barat. Sistem klasifikasi pada penelitian ini pun dapat membantu meningkatkan derajat kesehatan di lingkungan masyarakat Jawa Barat, akan tetapi terbatasnya ruang lingkup untuk memilih daerah yang akan diberikan bantuan kesehatan masih mengalami kendala dan dibutuhkannya suatu pemanfaatan data rekam medis untuk mengatasi kendala pada penelitian ini. Penelitian ini menggunakan sebuah Algoritma Iterative Dichotomiser 3 digunakan untuk mencari nilai layak dan tidaknya suatu daerah untuk diberikan bantuan kesehatan dan Metode Naïve Bayes Classifier digunakan untuk mencari tingkat prioritas bantuan kesehatan di masing-masing daerah dengan jenis penyakit mata. Proses pengelompokan dimulai dari menentukan jumlah class dari setiap field yang berada di data rekam medis untuk dihitung dan menghasilkan nilai gain. Hasil dari pengujian terhadap sistem yang dilakukan menghasilkan accuracy sebesar 87%.

Kata kunci: klasifikasi, penyakit, mata, Algoritma Iterative Dichotomiser 3, Naïve Bayes Classifier.

1. PENDAHULUAN

Sistem klasifikasi penyebaran penyakit mata merupakan suatu proses untuk mengetahui potensi penyakit mata di setiap daerah di Jawa Barat. Sistem klasifikasi pada penelitian ini pun dapat membantu meningkatkan derajat kesehatan di lingkungan masyarakat Jawa Barat, akan tetapi terbatasnya ruang lingkup untuk memilih daerah yang akan diberikan bantuan kesehatan masih mengalami kendala, hal ini dikarenakan kurangnya informasi untuk menentukan daerah dan pada penelitian terdahulu dalam klasifikasi kecenderungan penyakit mata menggunakan Association Rule dan Naïve Bayes Classifier, masih kurang akuratnya hasil yang diraih oleh sistem dengan penilaian terbesar ialah 51,31% untuk satu penyakit (Budiman, 2017), sehingga diperlukannya suatu analisis untuk dapat memberikan informasi dengan menganalisis data rekam medis dan data laporan Penanggulangan Gangguan Penglihatan dan Kebutaan (PGPK) yang kurang dimanfaatkan dan pada penelitian ini menggunakan Algoritma Iterative Dischotomiser 3 dan Metode Naïve Bayes Classifier.

Algoritma ID3 atau Iterative Dichotomiser 3 merupakan sebuah metode yang digunakan untuk membuat pohon keputusan. *Input* dari algoritma ini adalah sebuah *database* dengan beberapa variabel yang juga dikenal dengan atribut, setiap masukan dalam *database* menyajikan sebuah obyek dari domain yang disebut dengan variabel bebas, sebuah atribut dirancang untuk mengklasifikasikan obyek yang disebut dengan variabel tidak bebas (Shaleh, 2015) dan *Naïve Bayes Classifier* merupakan pengklasifikasian probabilitas sederhana berdasarkan pada teorema Bayes. Teorema Bayes dikombinasikan dengan “*Naïve*” yang berarti setiap atribut atau variabel bersifat bebas (*independent*). *Naïve Bayes Classifier* dapat dilatih dengan efisien dalam pembelajaran terawasi, proses *Naïve Bayes Classifier* mengasumsikan bahwa ada atau tidaknya suatu fitur pada suatu kelas tidak berhubungan dengan ada atau tidaknya fitur lain di kelas yang sama (Hamzah, 2012).

Tujuan yang dicapai pada penelitian ini yaitu mampu membuat sistem dengan membangun perangkat lunak dalam membantu rumah sakit dalam menentukan layak dan tidaknya daerah untuk pemilihan tempat kegiatan bantuan kesehatan dengan demikian daerah yang utama untuk dilakukan bantuan kesehatan dapat ditentukan dengan menggunakan Algoritma ID3 dan Naïve Bayes Classifier.

2. METODE PENELITIAN

Metode penelitian merupakan suatu tahapan proses penelitian yang digunakan untuk memecahkan persoalan yang muncul, metode penelitian didasarkan pada teori-teori yang mendukung pemecahan masalah dan bertujuan untuk memberikan gambaran pada penelitian yang dilakukan.

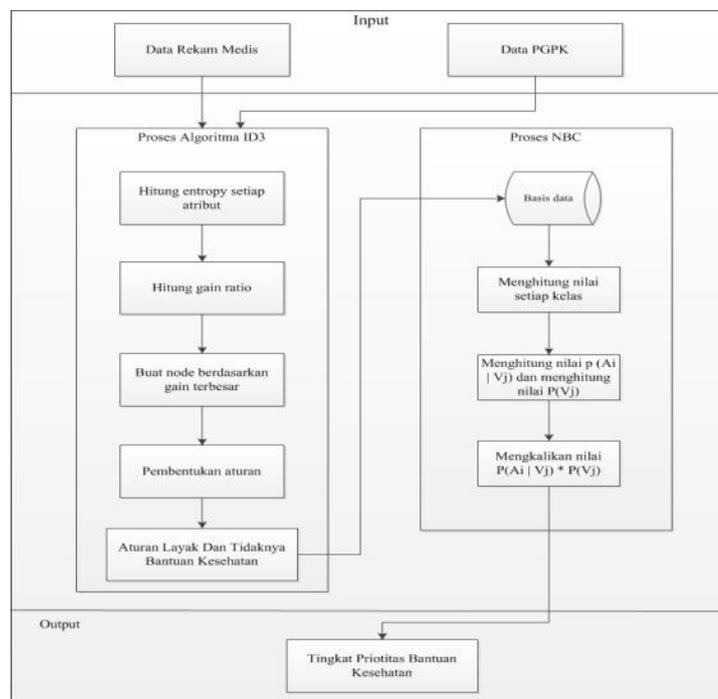
Adapun tahapan-tahapan metode penelitian yang dilakukan adalah:

a. Pengumpulan Data

Tahap dan cara pengumpulan data yang dilakukan dalam penelitian ini dengan cara teknik wawancara secara tatap muka dan tanya jawab langsung antara pengumpul data maupun peneliti terhadap narasumber atau sumber data dengan tujuan untuk memperoleh informasi.

b. Metode yang Digunakan

Metode yang digunakan pada penelitian ini yaitu Algoritma ID3 dilakukan untuk menentukan layak dan tidaknya daerah untuk dilakukan kegiatan bantuan kesehatan dan Metode Naïve Bayes Classifier yang digunakan untuk mengelompokkan tingkat penyakit mata dari berbagai kota dan kabupaten di Jawa Barat, gambaran umum sistem yang dibangun dapat dilihat pada Gambar 1.



Gambar 1. Gambaran umum sistem yang dibangun

Penjelasan gambaran umum sistem sebagai berikut:

1. Masukan (*Input*)

Masukan sistem yang akan dibangun pada penelitian ini berupa data rekam medis dan laporan data PGPK.

2. Proses (*Process*)

Sistem melakukan perhitungan dengan menerapkan Algoritma ID3 pada data rekam medis, tahapan pertama yaitu menghitung nilai *entropy* untuk setiap atribut, kedua menghitung nilai *gain ratio*, setelah menghitung nilai *entropy* dan nilai *gain ratio* selanjutnya membuat *node* dari nilai *gain* yang terbesar, setelah membuat *node* lalu pembuatan aturan atau *rule* untuk layak dan tidaknya bantuan kesehatan di suatu daerah dan ketika pembuatan aturan selesai maka akan dihitung dengan Naïve Bayes Classifier untuk mencari nilai tertinggi, dan nilai tertinggi tersebut akan menjadi acuan untuk melakukan prioritas daerah terhadap bantuan kesehatan dari hasil layak yang telah ditentukan oleh Algoritma ID3.

3. Keluaran (*Output*)

Keluaran atau *output* pada sistem ini adalah layak dan tidaknya suatu daerah untuk kegiatan bantuan kesehatan dan pengelompokan penyakit mata dengan tingkatan penyakit mata sesuai dengan kota dan kabupaten di Jawa Barat.

c. Metode Pengembangan Perangkat Lunak

Metode pengembangan perangkat lunak ini menggunakan Model Waterfall yang merupakan model untuk mengusulkan sebuah pendekatan kepada pengembangan perangkat lunak yang sistematis dan sekuensial yang meliputi aktivitas-aktivitas seperti analisis kebutuhan, desain, pengkodean, pengujian dan evaluasi, serta pemeliharaan.

3. HASIL DAN PEMBAHASAN

Proses klasifikasi dilakukan dari *node* paling atas yaitu akar pohon (*root*) yang dilanjutkan ke bawah melalui cabang-cabang sampai dihasilkan *node* daun (*leafes*) dimana *node* daun ini menunjukkan hasil akhir klasifikasi, sebuah obyek yang diklasifikasi dalam pohon harus diuji nilai *entropy* dan *ratio gain*.

3.1. Perhitungan Algoritma Iterative Dichotomiser 3

Langkah awal dalam menentukan layak dan tidaknya pemberian bantuan kesehatan penyakit mata ialah dengan menghitung nilai *entropy* untuk semua sampel (data latih) pada semua *node* kemudian pilih fitur dengan informasi *gain* yang maksimal dan gunakan fitur tersebut sebagai *node* pemecahan menjadi cabang. *Entropy* dapat dihitung menggunakan persamaan berikut:

$$E(s) = - \sum_{i=1}^m p_+(\omega_i | s) \log_2 p_-(\omega_i | s) \dots\dots\dots (1)$$

Langkah kedua *ratio gain* digunakan untuk memperkirakan pemilihan fitur yang tepat untuk menjadi pemecah pada *node* tersebut. *Gain* sebuah fitur ke-j dihitung menggunakan persamaan berikut:

$$G(s, j) = E(s) - \sum_{i=1}^m p(v_i | s) \times E(s_i) \dots\dots\dots (2)$$

3.1.1. Perhitungan Entropy

Langkah pertama ialah dimulai dari *node* akar, harus dihitung terlebih dahulu nilai *entropy* untuk *node* akar terhadap komposisi kelas dengan menggunakan persamaan 1.

$$\begin{aligned} E(\text{Semua}) &= -(p(\text{layak} | \text{semua}) \times \log_2 p(\text{layak} | \text{semua})) \\ &\quad + (p(\text{tidak} | \text{semua}) \times \log_2 p(\text{tidak} | \text{semua})) \\ &= -\left(\left(\frac{290}{2444} \times \log_2 \frac{290}{2444} \right) + \left(\frac{2154}{2444} \times \log_2 \frac{2154}{2444} \right) \right) \\ &= 0,10984 + 0,04834 \\ &= 0,15818 \end{aligned}$$

Perhitungan *entropy* dihitung hingga semua fitur kelas terhadap data selesai.

3.1.2. Perhitungan Ratio Gain

Langkah kedua ialah menghitung nilai *gain ratio* untuk semua fitur dengan menggunakan persamaan 2.

$$G(\text{semua}, \text{Kelamin}) = E(\text{semua}) - \sum_{i=1}^m p(v_i | \text{semua}) \times E(\text{semua}_{\text{Kelamin}})$$

$$\begin{aligned}
 &= E(semua) - \left((p((Perempuan | semua) \times E(semua_{Perempuan})) \right. \\
 &\quad \left. + (p((Laki - Laki | semua) \times E(semua_{Laki-Laki})) \right) \\
 &= 0,15818 - \left(\left(\frac{1485}{2444} \times 0,15629 \right) + \left(\frac{959}{2444} \times 0,16107 \right) \right) \\
 &= 0,15818 - ((0,09496) + (0,06320))
 \end{aligned}$$

dan seterusnya.

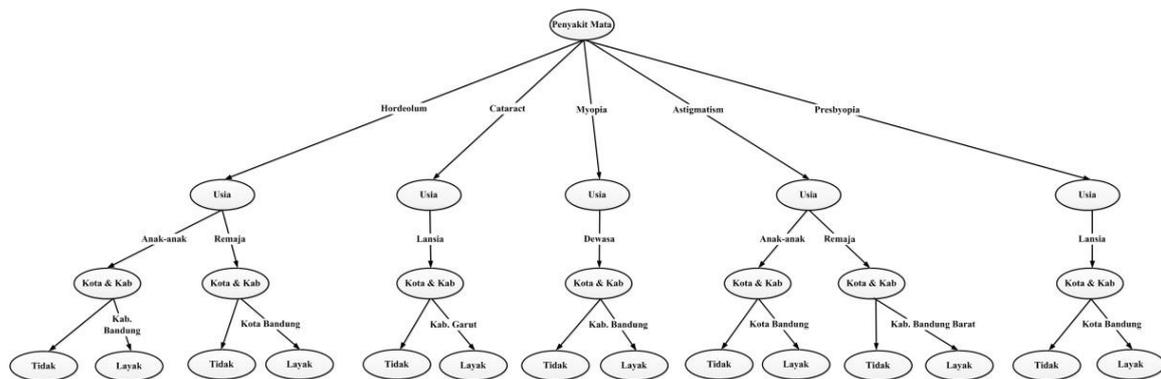
$G(semua, Kota dan Kabupaten) = 0,0231$

$G(semua, Usia) = 0,0062$

$G(semua, Jenis Diagnosa) = 0,0006$

$G(semua, Penyakit Mata) = 0,0354$

Hasil nilai *gain* tertinggi berdasarkan hasil perhitungan tersebut ialah fitur ‘penyakit mata’ sehingga dijadikan *node* akar, untuk cabangnya digunakan 5 nilai didalam *node* akar. Selanjutnya *node* akan terpecah menjadi 5 cabang atau 5 kelompok yaitu: Hordeolum, Cataract, Myopia, Astigmatism, dan Presbyopia, seperti pada Gambar 2.



Gambar 2. Klasifikasi Iterative Dichotomiser 3

3.2. Perhitungan Naïve Bayes Classifier

Tahapan ini yaitu proses untuk mencari penyakit mata yang tertinggi pada setiap penyakit mata yang bernilai ‘layak’ yang didapatkan dari hasil Algoritma ID3, pada langkah-langkah tahapan ini data yang akan diuji ialah data tabel pembentukan aturan. Naïve Bayes Classifier merupakan sebuah metode klasifikasi yang berakar pada teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu mencari peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes (Khadafy, 2015). Ciri utama dari Naïve Bayes Classifier ialah asumsi yang sangat kuat (naif) akan independensi dari masing-masing kondisi atau kejadian (Chandra, 2016 & I Gede, 2106).

$$P(A|B) = (P(B|A) * P(A))/P(B).....(3)$$

Langkah pertama yaitu membuat keterangan penyakit mata, kota serta kabupaten untuk mempermudah langkah dalam menentukan nilai *nc* untuk setiap *class* yang bernilai ‘layak’, dengan keterangan perhitungan pertama dalam keterangan penyakit mata, sebagai berikut:

Keterangan nilai *nc* Layak, Kota Bandung, Anak-anak:

- a. Layak.
- b. Kota Bandung.
- c. Anak-anak.

<p>1. Penyakit AAC $n = 3$ $p = 1/7 = 0,14285$ $\Sigma = 104$ a. $nc = 0$ b. $nc = 1$ c. $nc = 1$</p>	<p>2. Penyakit Astigmatism $n = 3$ $p = 1/7 = 0,14285$ $\Sigma = 104$ a. $nc = 1$ b. $nc = 1$ c. $nc = 1$</p>	<p>3. Penyakit Blepharitis $n = 3$ $p = 1/7 = 0,14285$ $\Sigma = 104$ a. $nc = 0$ b. $nc = 1$ c. $nc = 1$</p>
---	---	---

Langkah dalam tahapan menentukan nilai nc , *class* yang dipilih berjumlah tiga yaitu layak, Kota Bandung dan anak-anak dengan simbol yang ada pada tahapan di atas berupa n . Simbol p merupakan penyakit mata yang ada pada data sampel, tahapan ini penyakit mata yang diuji berjumlah satu dari delapan penyakit mata dan simbol Σ berupa jumlah pasien yang ada pada *class* Kota Bandung dengan yang bernilai layak.

Proses selanjutnya menghitung nilai $p(A_i|V_j)$ dan menghitung nilai $p(V_j)$ dalam langkah ini proses yang akan ditentukan berdasarkan langkah pertama yang telah diraih prosesnya. Langkah ke dua dalam menghitung nilai $p(A_i|V_j)$ dan menghitung $p(V_j)$.

Penyakit Acute Atopic Conjunctivitis (AAC)

- a. $P(\text{Acute Atopic Conjunctivitis}) = 0,13884$
- b. $P(\text{Acute Atopic Conjunctivitis}) = 0,14819$
- c. $P(\text{Acute Atopic Conjunctivitis}) = 0,14819$
- d. $P(\text{Acute Atopic Conjunctivitis}) = 0,14285 \dots$ dan seterusnya.

Langkah selanjutnya menghitung nilai $P(A_i|V_j) * P(V_j)$ untuk setiap v , langkah ini berfungsi untuk menentukan nilai tertinggi dari masing-masing *class* yang bertujuan untuk mencari nilai paling tinggi. Langkah selanjutnya untuk menghitung nilai $P(A_i|V_j) * P(V_j)$ untuk setiap v .

- 1. Penyakit AAC : $0,14285 * 0,13884 * 0,14819 * 0,14819 = 0,0004355446$
- 2. Penyakit Astigmatism : $0,14285 * 0,14819 * 0,14819 * 0,14819 = 0,0004648758$
- 3. Penyakit Blepharitis : $0,14285 * 0,13884 * 0,14819 * 0,14819 = 0,0004355446$
- 4. Penyakit CVD : $0,14285 * 0,13884 * 0,14819 * 0,14819 = 0,0004355446$
- 5. Penyakit Hordeolum : $0,14285 * 0,13884 * 0,14819 * 0,14819 = 0,0004355446$
- 6. Penyakit Myopia : $0,14285 * 0,13884 * 0,14819 * 0,14819 = 0,0004355446$

Setelah mengalikan nilai $P(A_i|V_j) * P(V_j)$ untuk setiap v , maka didapat nilai tertinggi yaitu penyakit Astigmatism, dan Kota Bandung dengan nilai sebesar 0,0004648758. Selanjutnya menghitung nilai nc dari nilai layak yang lain hingga semua nilai layak dihitung dengan Naïve Bayes Classifier dan didapat hasil klasifikasi seperti Tabel 1.

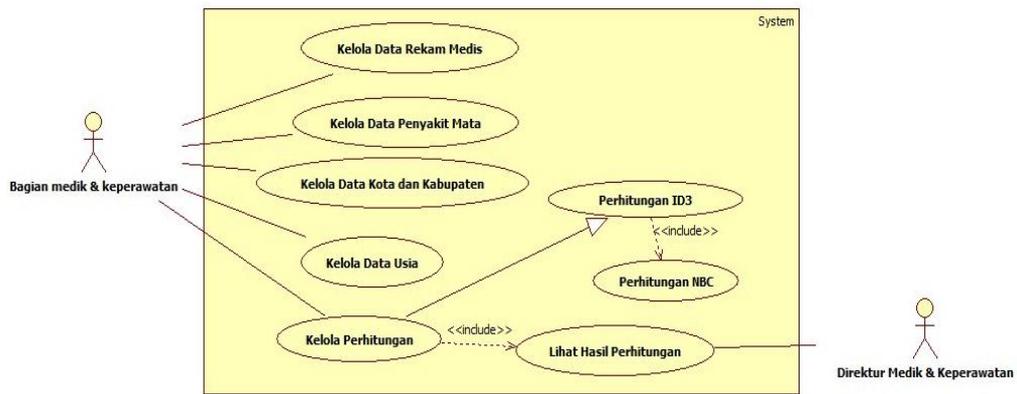
Tabel 1. Hasil Klasifikasi Naïve Bayes Classifier

Klasifikasi	Hasil
Kota Bandung, Anak-anak dan Astigmatism	0,0004648758
Kota Bandung, Anak-anak dan AAC	0,0004355446
Kota Bandung, Anak-anak dan Blepharitis	0,0004355446
Kota Bandung, Anak-anak dan CVD	0,0004355446
Kota Bandung, Anak-anak dan Hordeolum	0,0004355446
Kota Bandung, Anak-anak dan Myopia	0,0004355446

3.3. Perancangan Sistem

Perancangan sistem merupakan diagram yang menggambarkan fungsi yang dapat dijalankan dalam sebuah sistem terhadap aktor, interaksi antara aktor terhadap sistem akan digambarkan pada Use Case Diagram. Sistem yang akan dibangun terdapat dua aktor yang akan berinteraksi dengan sistem yaitu Bagian Medik dan Keperawatan dan Direktur Medik dan

Keperawatan, aktor-aktor tersebut dapat berinteraksi dengan sistem melalui fungsi-fungsi yang ada adalah seperti pada Gambar 3.



Gambar 3. Klasifikasi Penyakit Mata

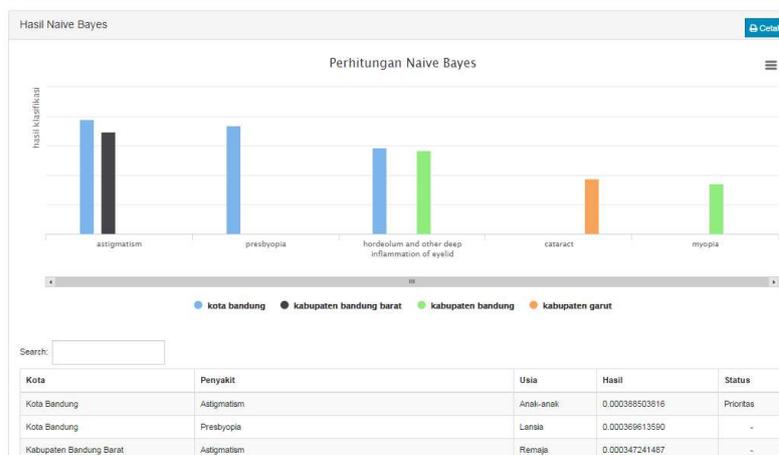
3.4. Hasil Penelitian

Hasil dari penelitian ini telah diujikan dengan menggunakan data rekam medis serta menggunakan Algoritma ID3 dan Metode Naïve Bayes Classifier. Hasil dari penelitian ini dapat dilihat pada Gambar 4 merupakan hasil perhitungan Algoritma ID3 dan Gambar 5 adalah grafik perhitungan Naïve Bayes Classifier.

Lihat Rule

No	If			Then
1	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kabupaten Bandung	Layak
2	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kabupaten Garut	Tidak Layak
3	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kabupaten Ciamis	Tidak Layak
4	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kabupaten Purwakarta	Tidak Layak
5	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kabupaten Bandung Barat	Tidak Layak
6	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kota Bandung	Tidak Layak
7	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kota Cimahi	Tidak Layak
8	Hordeolum And Other Deep Inflammation Of Eyelid	Anak-anak	Kota Tasikmalaya	Tidak Layak
9	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kota Bandung	Layak
10	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Sukabumi	Tidak Layak
11	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Cianjur	Tidak Layak
12	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Bandung	Tidak Layak
13	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Garut	Tidak Layak
14	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Ciamis	Tidak Layak
15	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Subang	Tidak Layak
16	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kabupaten Bandung Barat	Tidak Layak
17	Hordeolum And Other Deep Inflammation Of Eyelid	Remaja	Kota Cimahi	Tidak Layak

Gambar 4. Hasil sistem Algoritma ID3



Gambar 5. Hasil sistem Naïve Bayes Classifier

3.5. Pengujian Sistem

Berdasarkan hasil pengujian didapat 63% untuk *precision* dan 63% untuk *recall* yang menjelaskan tingkat ketepatan informasi yang diminta oleh Bagian Medik & Keperawatan yang diberikan oleh sistem, dan 87% untuk *accuracy* yang menentukan tingkat kedekatan dengan perhitungan manual dan perhitungan sistem.

4. KESIMPULAN

Kesimpulan pada penelitian ini telah dihasilkan sebuah sistem klasifikasi penyebaran penyakit mata di Jawa Barat dengan 8 (delapan) *usecase* dalam sistem menggunakan Algoritma ID3 dan Metode Naïve Bayes Classifier yang dapat membantu rumah sakit untuk kegiatan bantuan kesehatan dalam menentukan layak dan tidaknya bantuan di suatu daerah. Proses pada penelitian ini menggunakan data rekam medis dan data PGPK yang diperoleh dari Rumah Sakit Cicendo, Bandung Jawa Barat.

1. DAFTAR PUSAKA

- Budiman, T. Z., Witanti, W., Nursantika, D., (2017), Klasifikasi Penyebaran Penyakit Mata di Jawa Barat dengan Association Rule dan Naive Bayes Classifier. *Seminar Nasional Teknologi Informasi dan Multimedia*, ISSN 2302-3805.
- Chandra, D. N., Indrawan, G., (2016), Klasifikasi Berita Lokal Malang Menggunakan Metode Naive Bayes dengan Fitur N-Gram. *Jurnal Ilmiah Teknologi dan Informasi ASIA (JITIKA)*, vol. 10 no. 1, ISSN 0852-730X.
- Hamzah A., (2015), Klasifikasi Teks dengan Naive Bayes Classifier (NBC) untuk Pengelompokan Teks Berita dan Abstract Akademis, *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNA eST) Periode III*, ISSN 1979-911X.
- I Gede S. R., (2016), Identifikasi Jenis Obat Berdasarkan Gambar Logo pada Kemasan Menggunakan Metode Naive Bayes. *Jurnal Sisfo*, vol. 06 no. 01, pp. 17-32.
- Khadafy, A. R., (2015), Penerapan Naive Bayes untuk Mengurangi Data Noise pada Klasifikasi Multi Kelas dengan Decision Tree. *Journal of Intelligent Systems*, vol. 1 no. 2, ISSN 2356-3982.
- Safuan, R. C., (2015), Penanganan Fitur Kontinyu dengan Feature Discretization Berbasis Expectation Maximization Clustering untuk Klasifikasi Spam Email Menggunakan Algoritma ID3. *Journal of Intelligent Systems*, vol. 1 no. 2, ISSN 2356-3982.
- Saleh, A., (2015), Implementasi Metode Klasifikasi Naive Bayes dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga, *Citec Journal*, vol. 2 no. 3, ISSN 2354-5771.
- Selly, A. Z., (2014), Penggunaan Pohon Keputusan untuk Klasifikasi Tingkat Kualitas Mahasiswa Berdasarkan Jalur Masuk Kuliah. *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*, ISSN 1907-5022.
- Yusuf, E. P., (2012), Pemilihan Kriteria Splitting dalam Algoritma Iterative Dichotomiser 3 untuk Penentuan Kualitas Beras Studi Kasus Perum Bulog Divre Lampung. *Jurnal TELEMATIKA MKOM*, vol 4 no. 1, ISSN 2085-725X.