

IMPLEMENTASI ALGORITMA C4.5 UNTUK KLASIFIKASI PENYAKIT INFEKSI SALURAN KEMIH BERBASIS WEB

¹Muhammad Fikri Hasani, ²Sutikno

¹Computer Science Department, School of Computer Science, Bina Nusantara University, ²Program Studi Teknik Informatika, Fakultas Sains dan Matematika, Universitas Diponegoro.

¹Jl. Raya Kb. Jeruk No.27, Kota Jakarta Barat 11530, Daerah Khusus Ibukota Jakarta, ²Jl. Prof. Soedarto No.50275, Tembalang, Kec. Tembalang, Kota Semarang, Jawa Tengah

¹muhammad.fikri003@binus.ac.id, ²sutikno@lecturer.undip.ac.id

Abstrak

Infeksi saluran kemih merupakan penyakit sistem saluran kemih yang disebabkan oleh bakteri, dengan 90-100 kasus per 100.000 penduduk pertahun. Infeksi saluran kemih terdiri dari infeksi saluran kemih bawah atau cystitis dan infeksi saluran kemih atas atau pyelonephritis. Kedua penyakit ini memiliki ciri-ciri seperti sering buang air, nyeri punggung bawah, mual-mual, rasa perih saat berkemih, atau rasa terbakar pada alat kelamin. Sistem yang dibuat bertujuan untuk membantu dokter dalam melakukan diagnosa awal kepada pasien berdasarkan gejala yang dimiliki pasien, apakah menderita cystitis, pyelonephritis, keduanya, atau tidak menderita kedua penyakit tersebut. Metode pohon keputusan C4.5 digunakan pada penelitian ini karena diketahui memiliki tingkat akurasi yang tinggi pada penelitian-penelitian sebelumnya yang didominasi parameter data nomimal. Pengujian kinerja sistem yang dibangun menggunakan pengujian akurasi, sensitivity dan specificity. Hasil pengujian menghasilkan akurasi rata-rata 95%, sensitivity rata-rata sebesar 96.667%, dan specificity rata-rata sebesar 92.226% pada model pohon cystitis, diikuti oleh model pohon pyelonephritis dengan akurasi rata-rata sebesar 94%, sensitivity rata-rata sebesar 90%, dan specificity rata-rata sebesar 97.143%. Selain itu, pengujian menggunakan blackbox texting terhadap requirements sistem juga menunjukkan bahwa sistem dapat dijalankan dengan baik.

Kata Kunci: C4.5, cystitis, pyelonephritis

Abstract

Urinary tract infection is a disease of the urinary system caused by bacteria, with 90-100 cases per 100,000 population per year. Urinary tract infections consist of lower urinary tract infection or cystitis and upper urinary tract infection or pyelonephritis. Both diseases have characteristics such as frequent urination, lower back pain, nausea, burning sensation when urinating, or burning sensation in the genitals. The system created aims to assist doctors in making initial diagnoses to patients based on the symptoms the patient has, whether they suffer from cystitis, pyelonephritis, both, or not suffering from both diseases. The C4.5 decision tree method was used in this study because it is known to have a high level of accuracy in with majority number of nominal data. Testing the performance of the system built using accuracy, sensitivity, and specificity testing. The test results produce an average accuracy of 95%, an average sensitivity of 96.667%, and an average specificity of 92.226% on the cystitis tree model, followed by the pyelonephritis tree model with an average accuracy of 94%, an average sensitivity of 90%, and the average specificity is 97.143%. In addition, testing using blackbox texting on system requirements also shows that the system can run well.

Keywords: C4.5, cystitis, pyelonephritis

PENDAHULUAN

Penyakit infeksi saluran kemih merupakan penyakit yang menyerang saluran kemih, dari saluran kemih bawah sampai ginjal yang disebabkan oleh bakteri [1]. Penyakit saluran kemih memiliki jumlah pasien yang lumayan banyak dinomi Indonesia yaitu pada tahun 2014 berjumlah 90-100 kasus per 100.000 orang per tahun [2]. Penyakit infeksi saluran kemih sendiri terbagi menjadi 2, yaitu infeksi saluran kemih bawah atau *cystitis*, dan infeksi saluran kemih atas atau *pyelonephritis* [3,4]. *Cystitis* merupakan penyakit peradangan kantung kemih sedangkan *pyelonephritis* adalah nephritis pada *renal pelvis*. Gejala-gejala yang dimiliki oleh kedua penyakit ini diantaranya adalah sering buang air, nyeri punggung bawah, mual-mual, rasa perih saat berkemih, atau rasa terbakar pada alat kelamin [3,4]. Gejala-gejala ini, kerap kali muncul pada penyakit-penyakit infeksi saluran kemih dan tidak spesifik salah satu penyakit saja. Berdasarkan kasus yang diangkat, dibutuhkan sebuah sistem yang dapat melakukan klasifikasi penyakit penyakit infeksi saluran kemih yang diderita oleh pasien berdasarkan pada gejala-gejala yang dialami pasien sehingga dapat membantu dokter dalam melakukan diagnosis.

Diagnosis merupakan tindakan untuk menentukan kondisi kesehatan yang sedang dialami oleh seseorang sebagai dasar pengambilan keputusan medis [5]. Proses diagnosis adalah proses yang penting untuk

dilakukan terhadap pasien, untuk mengetahui penyakit apa yang diderita, perkembangan dari penyakit tersebut, dan tindakan medis yang akan dilakukan oleh seorang dokter. Seiring berjalannya waktu, kecerdasan buatan semakin digunakan untuk membantu diagnosis penyakit yang dialami oleh pasien seorang dokter, salah satu nya adalah melakukan identifikasi terhadap penyakit yang diderita oleh pasien. Identifikasi penyakit ini dapat dilakukan oleh salah satu cabang dari kecerdasan buatan yaitu *supervised learning*. *Supervised learning* adalah sebuah cabang dari kecerdasan buatan yang akan melakukan estimasi prediksi keluaran berdasarkan data masukan [6]. Proses estimasi bisa berbentuk sebuah fungsi yang akan memberikan keluaran berdasarkan data masukan, atau bisa juga berbentuk seperti pohon keputusan. *Supervised learning* mendapatkan namanya dari proses yang dilakukan, yaitu setiap algoritma akan berlatih dengan data yang disebut data latih berdasarkan pada label yang dimiliki oleh setiap data latih, atau bisa dianggap bahwa proses pelatihan akan disupervisi oleh label tersebut. Data yang dikandung oleh label yang digunakan untuk algoritma *supervised learning* disebut sebagai target. Jika target merupakan data nominal, maka data tersebut juga disebut sebagai kelas. Kelas inilah yang digunakan pada penelitian identifikasi penyakit ataupun klasifikasi penyakit oleh beberapa algoritma. Penelitian-penelitian yang bertujuan untuk membantu proses identifikasi penyakit dengan algoritma-

algoritma *supervised learning* telah banyak dilakukan. *Support Vector Machine* digunakan untuk melakukan identifikasi penyakit kanker payudara [7] dan klasifikasi penyakit saluran pernapasan [8]. Algoritma *naïve bayes* digunakan untuk melakukan identifikasi penyakit diabetes [9] dan klasifikasi penyakit gigi [10]. Algoritma C4.5 digunakan untuk proses identifikasi penyakit *pneumonia* [11], serta algoritma *artificial neural network* digunakan untuk melakukan identifikasi penyakit kanker paru-paru [12].

Seperti yang sudah dijelaskan sebelumnya, algoritma *supervised learning* memerlukan adanya data latih. Data latih yang digunakan untuk penelitian ini adalah *open data* dari *University of California Irvine* (UCI) Machine Learning Repository [13]. Data tersebut memiliki 120 data dengan 2 buah target kelas penyakit yaitu *cystitis* dan *pyelonephritis* dengan 6 buah parameter gejala penyakit. Data yang diambil kemudian dianalisa dan ditemukan bahwa hampir semua parameter data merupakan data nominal atau kategorikal kecuali parameter temperatur suhu pasien, sehingga digunakan algoritma C4.5 sebagai model kecerdasan buatan yang akan digunakan. Algoritma C4.5 sendiri adalah algoritma *supervised learning* untuk melakukan klasifikasi berdasarkan aturan-aturan yang terbentuk pada proses pelatihan. Aturan-aturan ini akan membentuk pohon keputusan dengan daun dari pohon yang terbentuk adalah prediksi model C4.5 [14]. Algoritma C4.5 memiliki tingkat keakuratan

yang tinggi jika parameter data yang digunakan mayoritas bernilai nominal seperti pada penelitian klasifikasi penyakit jantung koroner dengan nilai AUC 84.2% [15], klasifikasi penyakit *hepatitis* menggunakan metode C4.5 dengan hasil akurasi sebesar 85.81% [16], serta klasifikasi penyakit ikan dengan akurasi sebesar 88.4% [17]. Penelitian ini bertujuan untuk melakukan implementasi algoritma C4.5 pada sistem berbasis web untuk melakukan klasifikasi penyakit saluran kemih. Sistem yang dibangun dapat mengklasifikasikan penyakit infeksi saluran kemih berdasarkan pada gejala yang dialami oleh pasien.

METODE PENELITIAN

Metode penelitian yang digunakan di dalam penelitian ini terdiri dari beberapa tahapan, yaitu pengumpulan data, perancangan, implementasi model C4.5, implementasi sistem dan pengujian.

Pengumpulan Data

Pada tahapan ini, data dikumpulkan dan dianalisa. Data yang digunakan adalah *open data* yang diambil dari UCI Machine Learning Repository. *Dataset* memiliki 6 gejala pasien penyakit saluran kemih, 5 diantaranya merupakan data nominal/kategorikal dengan nilai *yes* dan *no*, dan 1 gejala bernilai kontinyu yaitu suhu pasien. Detail dari data yang digunakan dapat dilihat pada tabel 1. Selain 6 kolom sebagai parameter, *dataset* memiliki 2

kolom label yaitu *inflammation of urinary bladder* atau disebut *cystitis*, dan *nephritis of renal pelvis origin* atau disebut *pyelonephritis*. Kedua label tersebut dapat dilihat pada dua baris terakhir tabel 1. Kedua label ini memiliki kelas *yes* dan *no*. Selanjutnya, label *Inflammation of urinary bladder* akan dinamakan sebagai *cystitis* sesuai dengan nama medisnya, dan label *Nephritis of renal pelvis origin* akan dinamakan sebagai

pyelonephritis. Contoh sampel data yang dimiliki dapat dilihat pada tabel 2. Tabel 2 menunjukkan bahwa *dataset* memiliki beberapa karakteristik terhadap dua label yang dimiliki:

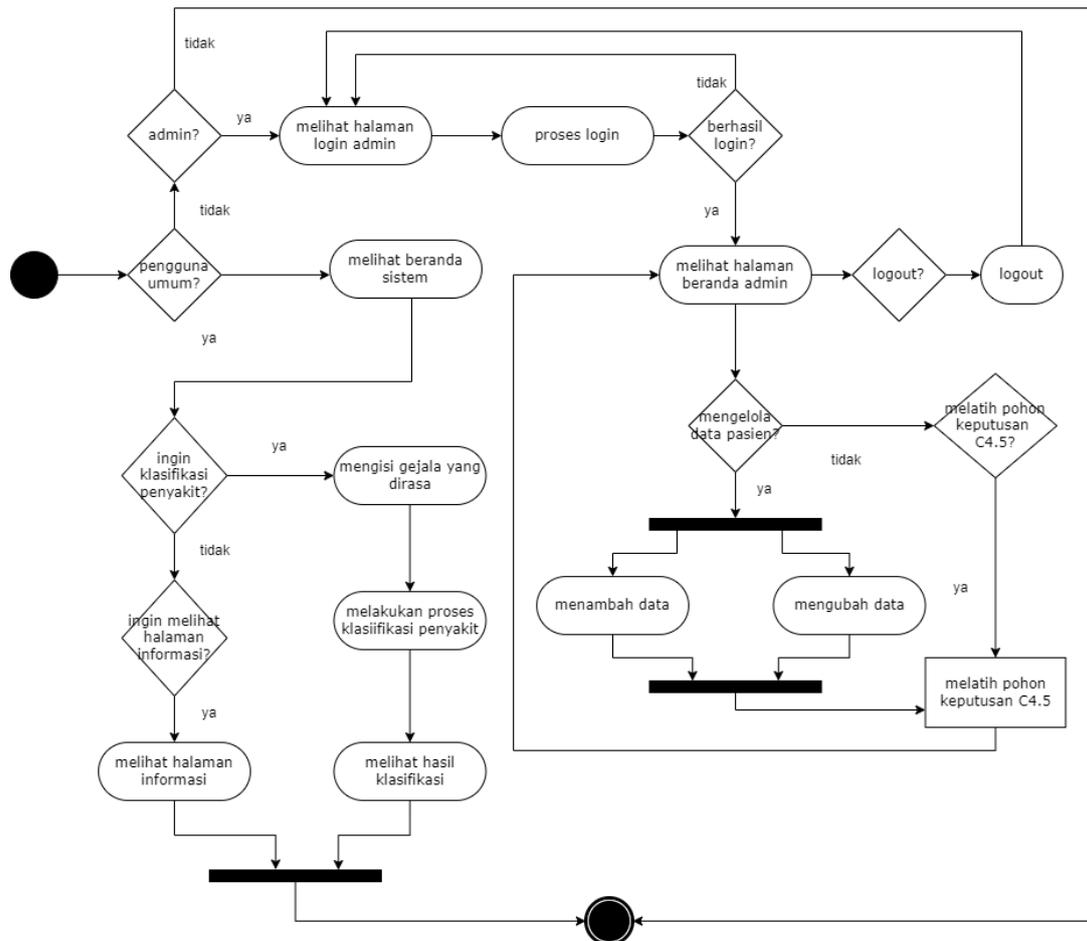
1. Sebuah data masuk ke salah satu dari *cystitis* atau *pyelonephritis*.
2. Sebuah data masuk ke kedua kelas *cystitis* dan *pyelonephritis*.
3. Sebuah data tidak masuk ke kedua penyakit

Tabel 1. Dataset Infeksi Kandung Kemih UCI Machine Learning Repository.

Kolom	Deskripsi
<i>Temperature</i>	Paramater. Temperatur suhu pasien, berkisaran dari 35C – 42C. Bernilai kontinyu.
<i>Nausea</i>	Parameter. Terjadi mual-mual dan keinginan untuk muntah [18]. Bernilai nominal dengan nilai <i>yes</i> dan <i>no</i> .
<i>Lumbar pain</i>	Paramater. Terjadi rasa nyeri pada punggung bawah [19,20]. Bernilai nominal <i>yes</i> dan <i>no</i> .
<i>Urine push</i>	Paramater. Frekuensi buang air yang tinggi. Bernilai nominal <i>yes</i> dan <i>no</i> .
<i>Mictur pain</i>	Parameter. Terjadi rasa nyeri ketika buang air kecil. Bernilai nominal <i>yes</i> dan <i>no</i> .
<i>Burning pain</i>	Parameter. Ada rasa terbakar pada alat kelaminr dengan nilai <i>yes</i> dan <i>no</i>
<i>Cystitis</i>	Label. Label penyakit <i>cystitis</i> atau <i>inflammation of urinary bladder</i> . Memiliki 61 jumlah kelas <i>yes</i> dan 59 jumlah kelas <i>no</i> .
<i>Pyelonephritis</i>	Label. Label penyakit <i>pyelonephritis</i> atau <i>nephritis of renal pelvis origin</i> . Memiliki 70 jumlah kelas <i>yes</i> dan 50 jumlah kelas <i>no</i> .

Tabel 2. Contoh Sampel Dataset.

<i>Temperature</i>	<i>Nausea</i>	<i>Lumbar Pain</i>	<i>Urine Push</i>	<i>Mictur Pain</i>	<i>Burning Pain</i>	<i>Cystitis</i>	<i>Pyelonephritis</i>
35,9	no	no	yes	yes	yes	yes	no
35,9	no	yes	no	no	no	no	no
36,0	no	no	yes	yes	yes	yes	no
40,6	yes	yes	no	yes	no	no	yes
40,7	yes	yes	yes	yes	yes	yes	yes



Gambar 1. Alur Kerja Sistem

Perancangan.

Pada tahapan ini, dilakukan analisa alur kerja sistem yang dibangun, kebutuhan sistem dan analisa terhadap data yang sudah dikumpulkan. Sistem ini memiliki dua buah akses pengguna yaitu *pengguna umum* dan *admin*. *Pengguna umum* adalah pengguna yang menggunakan sistem untuk mendapatkan hasil klasifikasi penyakit kantung kemih, sedangkan *admin* adalah pengguna yang diberikan akses untuk memperbaiki data latih dan melatih kembali model kecerdasan buatan. *Pengguna umum* disini dapat merupakan pasien yang menggunakan sistem saat berobat

di tempat. Pengguna *admin* adalah dokter. Tahapan perancangan ini juga melakukan *use case analysis* yang akan digunakan sebagai basis pengujian *black box*.

Implementasi Model Algoritma C4.5.

Tahapan implementasi model C4.5 merupakan tahapan pengembangan kecerdasan buatan C4.5. Implementasi dari model algoritma C4.5 menggunakan *Stratified K-Fold Cross Validation* untuk melihat apakah model algoritma C4.5 memiliki performa yang rata untuk seluruh bagian dari data latih. Proses membangun model algoritma C4.5

menggunakan konsep *divide and conquer* dimana terdiri dari 2 poin [14]:

1. Jika semua kasus pada data set S masuk kedalam kelas yang sama, misal C_j , maka pohon keputusan yang terbentuk memiliki daun dengan label C_j .
2. Jika tidak, asumsikan B adalah sebuah tes dengan keluaran b_1, b_2, \dots, b_i yang menghasilkan partisi dari T , dinotasikan dengan T_i , yaitu kumpulan kasus pada T yang memiliki hasil b_i dari B .

Proses pelatihan pohon keputusan akan melakukan iterasi pada poin kedua sampai ditemukan kondisi poin pertama. Setiap sub pohon akan dilakukan proses pemilihan kandidat akar sampai tidak ada sub pohon yang dapat dibentuk. Proses pemilihan kandidat akar, menggunakan 2 teori yaitu *information gain* dan *gain ratio* [14]. *Information gain* merupakan cara mengukur pengurangan ketidakmurnian informasi jika sebuah atribut X digunakan dalam melakukan partisi kasus T , dimana semakin besar nilai dari *information gain* maka semakin murni data yang dimiliki. Persamaan untuk mendapatkan nilai *information gain* dapat dilihat pada persamaan (4). Untuk menghitung *information gain*, diperlukan juga menghitung *entropy*. Persamaan untuk mendapatkan *entropy* dari atribut x dapat dilihat pada persamaan (3), dimana *entropy* (T) merupakan *entropy* dari semesta kasus, $Pr(C_j)$ adalah probabilitas kelas C ke j . $|T_i|$ merupakan jumlah kasus pada partisi data ke- i , dan T merupakan jumlah kasus dalam T . *Entropy* sendiri adalah cara

mengukur tingkat ketidakmurnian informasi dari sebuah data. Semakin kecil nilai sebuah *entropy*, berarti semakin murni informasi yang didapatkan dari sebuah atribut.

$$entropy(T) = -\left(\sum_{j=1}^k Pr(C_j) * \log_2(Pr(C_j))\right) \quad (1)$$

$$Pr(C_j) = \frac{freq(C_j, T)}{|T|} \quad (2)$$

$$entropy_X(T) = \sum_{i=1}^n \frac{|T_i|}{|T|} * entropy(T_i) \quad (3)$$

$$gain(X) = entropy(T) - entropy_X(T) \quad (4)$$

Setelah mendapatkan nilai *information gain* untuk parameter X , maka dapat dihitung nilai dari *gain ratio* parameter X . *Gain ratio* digunakan karena *information gain* memiliki kelemahan yaitu menghasilkan bias yang kuat jika kasus yang ditangani memiliki parameter dengan nilai yang banyak, kelas keluaran yang banyak, atau nilai data pada suatu parameter bernilai unik. Persamaan untuk mendapatkan *gain ratio* dapat dilihat pada persamaan 6 dengan sebelumnya menghitung terlebih dahulu nilai dari *split information* yang dapat dilihat pada persamaan 5.

$$split\ info(X) = -\left(\sum_{i=1}^n \frac{|T_i|}{|T|} * \log_2\left(\frac{|T_i|}{|T|}\right)\right) \quad (5)$$

$$gain\ ratio(X) = \frac{gain(X)}{split\ info(X)} \quad (6)$$

Proses perhitungan *gain ratio* dilakukan berulang untuk semua sisa parameter pada

kasus T. Kemudian dicari *gain ratio* tertinggi untuk menjadi akar. Pendekatan perhitungan diatas dilakukan pada data dengan parameter nominal yaitu *nausea*, *lumbar pain*, *burning pain*, *urine push*, dan *mictur pain*. Data kontinyu untuk kolom temperature tidak dapat langsung dilakukan *split* seperti data nominal, maka dilakukan proses yang berbeda, yaitu perhitungan *gain ratio* menggunakan data *median* dari seluruh data yang ada pada proses tersebut.

Model algoritma pertama kali akan dibangun dengan menggunakan *stratified k-fold cross validation* untuk melihat kemangkusan dari algoritma C4.5 terhadap data yang dimiliki. Nilai K yang digunakan sebesar 10, sehingga dari total 120 data, 12 data akan menjadi data uji dan 108 data akan menjadi data latih. Setiap *fold* akan dihitung akurasi, *sensitivity* dan *specificity*. *Sensitivity* adalah sebuah metrik untuk menentukan seseorang memiliki penyakit, dan *specificity* adalah sebuah metrik untuk menentukan seseorang tidak memiliki penyakit [21]. Semakin tinggi nilai *sensitivity* maka semakin tinggi performa model kecerdasan buatan dalam melakukan prediksi seseorang mengidap penyakit, dan semakin tinggi nilai *specificity* maka semakin tinggi performa model kecerdasan buatan dalam melakukan prediksi seseorang tidak mengidap penyakit. Setelah melakukan proses *k-fold cross validation*, pohon keputusan akan dibangun dengan seluruh data yang dimiliki. Berdasarkan pada tabel 1 dan tabel 2, model

algoritma pohon keputusan C4.5 akan dibentuk 2 buah yaitu untuk pohon keputusan *cystitis*, dan pohon keputusan *pyelonephritis*. Pembuatan 2 pohon keputusan dikarenakan data yang dimiliki memiliki 2 label penyakit yaitu *cystitis* dan *pyelonephritis*, serta setiap label memiliki kelas *yes* dan *no*. Sehingga, setiap pohon keputusan akan melakukan proses klasifikasi untuk setiap label.

Implementasi Sistem.

Tahapan implementasi sistem dilakukan pembuatan sistem berdasarkan hasil dari tahapan perancangan. Tahapan ini, sistem akan direalisasikan berdasarkan alur sistem pada gambar 1. Selain itu, pada tahapan ini sistem dibangun dengan bahasa pemrograman PHP dan framework Yii2.

Pengujian.

Tahapan pengujian akan melakukan pengujian terhadap model kecerdasan buatan untuk mendapatkan akurasi, *sensitivity* dan *specificity*, serta pengujian terhadap sistem dengan metode *black box testing*.

HASIL DAN PEMBAHASAN

Pertama-tama, dilakukan proses pelatihan model C4.5 sebagai otak dari sistem yang dibuat dengan K-fold cross validation seperti pada metode penelitian. Hasil perhitungan akurasi, *sensitivity* dan *specificity* dapat dilihat pada tabel 3 dan 4. Tabel 3 rata-rata akurasi bernilai 95%, *sensitivity* bernilai

96.667% dan specificity bernilai 92.226% untuk pohon keputusan C4.5 cystitis. Tabel 4 menunjukkan hasil perhitungan metrik pohon keputusan C4.5 pyelonephritis dengan rata-rata akurasi bernilai 94%, rata-rata sensitivity 90%, dan rata-rata specificity bernilai 97.143%. Angka ini berarti model algoritma C4.5 yang dibangun mampu melakukan klasifikasi bahwa seseorang memiliki penyakit

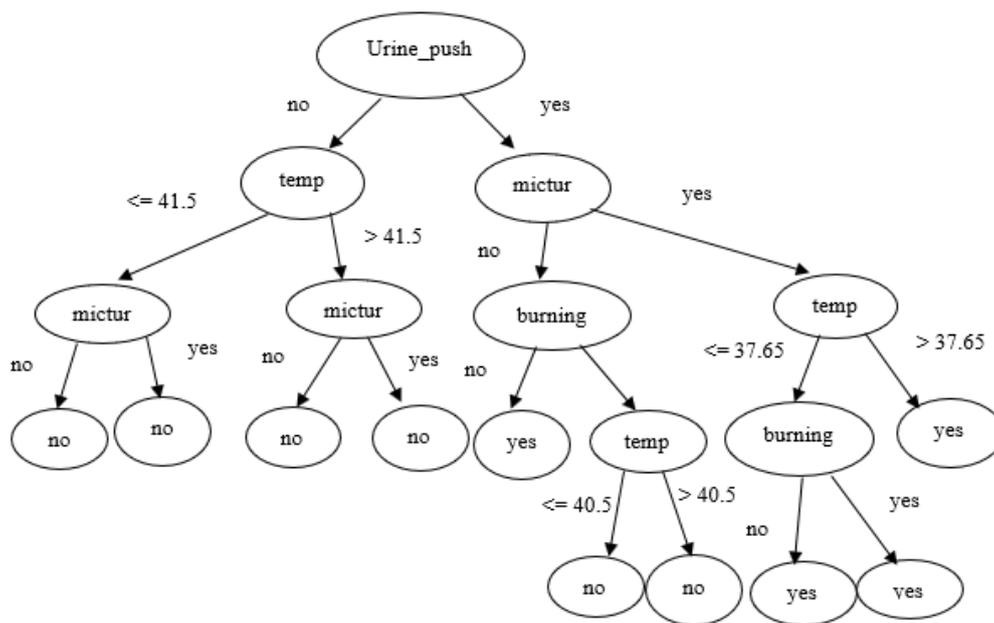
cystitis sebesar 96.667%, dan jika sistem melakukan klasifikasi seseorang tidak menderita cystitis, maka hasil klasifikasi tersebut kemungkinan benar sebesar 92.226%. Tahapan selanjutnya adalah melakukan implementasi C4.5 pada seluruh *dataset*. Gambar 2 adalah model pohon keputusan untuk *cystitis* dan gambar 3 adalah model pohon untuk *pyelonephritis*.

Tabel 3. Hasil *Stratified K-fold Cross Validation* terhadap Pohon Keputusan *Cystitis*

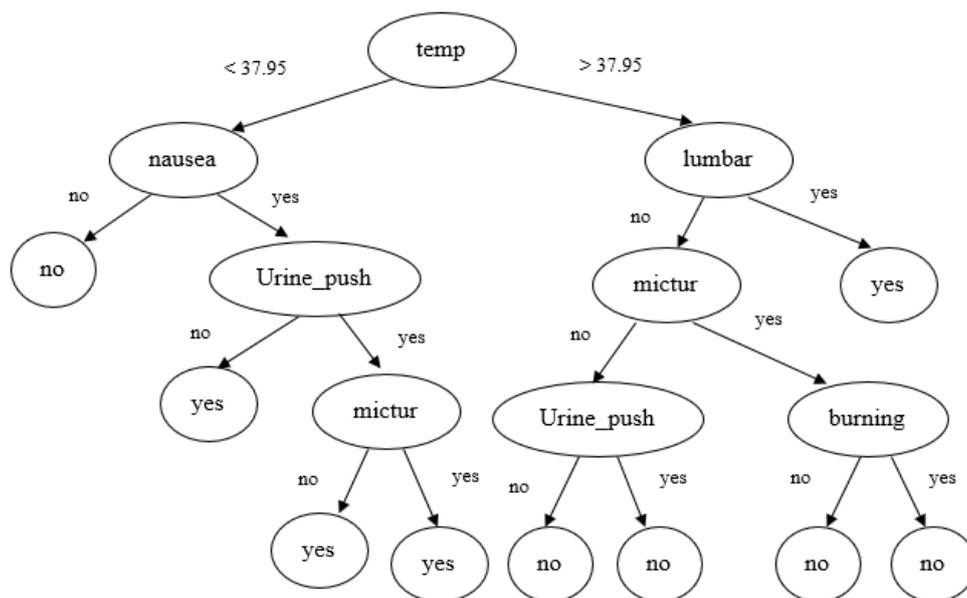
<i>Fold</i>	Hasil Pengujian Kinerja Pohon <i>Cystitis</i>		
	<i>Akurasi</i>	<i>Sensitivity</i>	<i>Specificity</i>
1	83.333	100	71.428
2	91.667	66.667	100
3	100	100	100
4	91.667	100	83.333
5	100	100	100
6	100	100	100
7	91.667	100	80
8	100	100	100
9	100	100	100
10	100	100	100
Rata-rata	95	96.667	92.226

Tabel 4. Hasil *Stratified K-fold Cross Validation* terhadap Model Pohon Keputusan *Pyelonephritis*

<i>Fold</i>	Hasil Pengujian Kinerja Pohon <i>Pyelonephritis</i>		
	<i>Akurasi</i>	<i>Sensitivity</i>	<i>Specificity</i>
1	91.667	100	85.715
2	91.667	100	85.715
3	91.667	80	100
4	91.667	80	100
5	100	100	100
6	83.333	60	100
7	100	100	100
8	91.667	80	100
9	100	100	100
10	100	100	100
Rata-rata	94	90	97.143



Gambar 2. Pohon Keputusan *Pyelonephritis*.



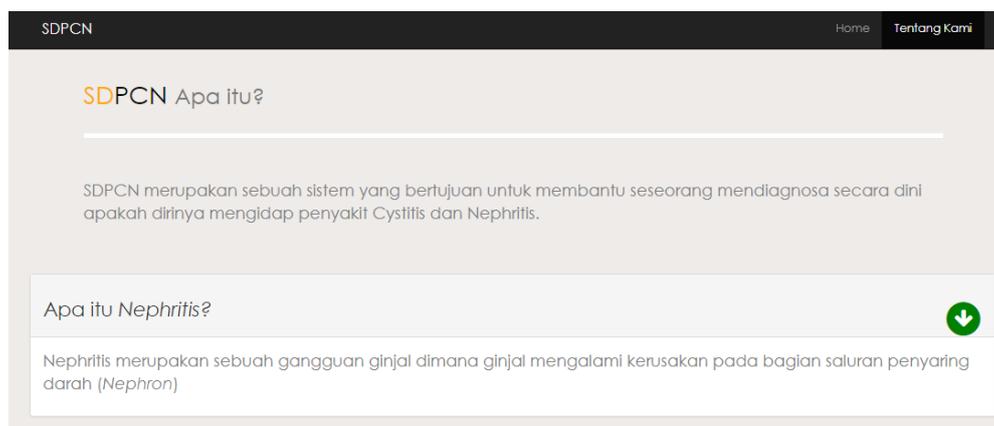
Gambar 3. Pohon Keputusan *Pyelonephritis*.

Kemudian, dibentuk tampilan antar muka sebagai implementasi dari analisa alur kerja sistem pada gambar 1. Gambar 4 – 7 merupakan gambar implementasi dari sistem yang dibuat. Gambar 4 merupakan contoh

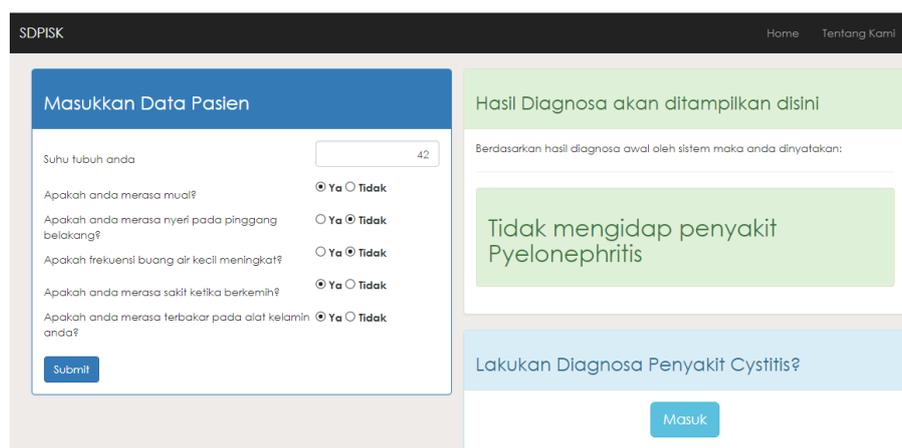
antar muka informasi seputar penyakit kantung kemih yang akan diklasifikasikan oleh sistem, yaitu *cystitis* dan *pyelonephritis*. Halaman ini memberikan informasi cara kerja dari sistem, deskripsi dari sistem tersebut, serta deskripsi

dari masing-masing penyakit. Pada Gambar 5 pengguna akan diminta untuk memasukkan gejala yang dimiliki, kemudian sistem akan menampilkan hasil klasifikasi. Pada halaman ini, model algoritma C4.5 yang telah dibentuk akan dibaca dari database menjadi sebuah *array associative* [22]. Parameter yang dimasukkan kemudian akan ditelusuri alurnya

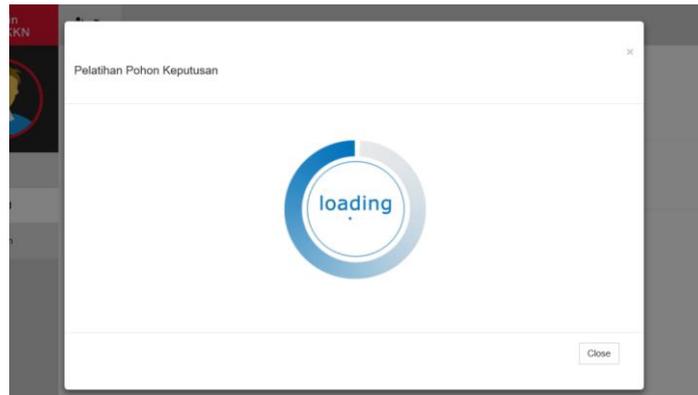
berdasarkan pohon keputusan yang telah dibuat. Hasil klasifikasi kemudian ditampilkan ke halaman. Gambar 6 merupakan antar muka proses pelatihan yang dilakukan oleh admin. Agar terlihat proses yang terjadi maka disediakan animasi *loading* agar. Data yang digunakan oleh sistem disimpan di dalam *database*.



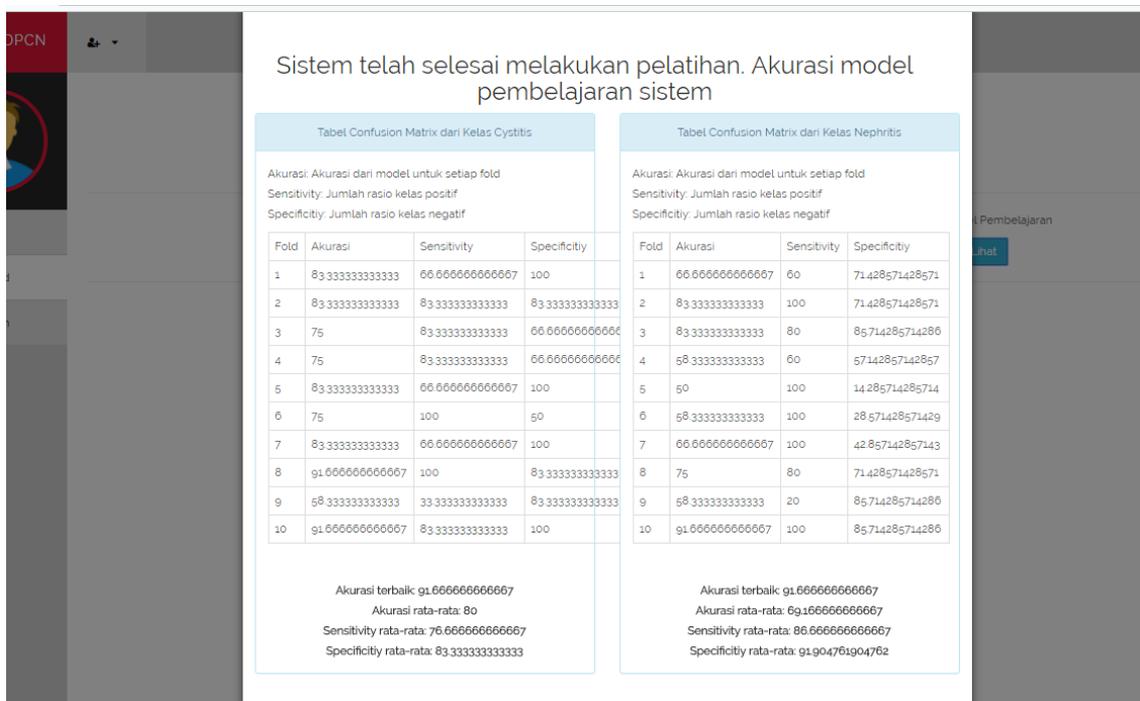
Gambar 4. Implementasi Antarmuka Melihat Halaman Informasi Sistem



Gambar 5. Implementasi Antarmuka Melakukan Diagnosa *Pyelonephritis*– Hasil Diagnosa



Gambar 6. Implementasi Antarmuka Melakukan Pelatihan – Proses Pelatihan



Gambar 7. Implementasi Antarmuka Melakukan Pelatihan – Hasil Pelatihan

Gambar 7 merupakan antar muka tampilan informasi yang muncul setelah melakukan training model algoritma C4.5. Pada tampilan ini akan ditampilkan nilai akurasi, *specificity* dan *sensitivity* setiap *fold*, serta rata-rata keseluruhan *fold*. Sistem kemudian diuji menggunakan *blackbox testing*. Semua skenario test seperti “Melihat

hasil diagnosa apakah dirinya mengidap penyakit *cystitis* atau *pyelonephritis*”, “Admin dapat melakukan pelatihan model kecerdasan buatan”, “Melihat halaman informasi” dapat dilakukan dengan baik. Namun, ditemukan kekurangan dari sistem. Sistem memiliki akses admin untuk memperbaharui data. Perubahan data ini dapat dalam bentuk penambahan data

baru atau modifikasi/update data yang sudah ada. Perubahan data yang dilakukan oleh admin dapat merubah komposisi jumlah data setiap penyakit, dan memberikan fenomena *data imbalance*, sehingga perlu ada mekanisme untuk mengatasi *data imbalance* pada sistem yang dibuat.

KESIMPULAN DAN SARAN

Algoritma C4.5 telah berhasil diimplementasikan menggunakan sistem berbasis web. Semua fitur yang diturunkan dari analisis alur sistem mampu dijalankan oleh pengguna, serta *metrics* dari model algoritma C4.5 memberikan hasil yang baik dimana rata-rata untuk akurasi, *specificity* dan *sensitivity* untuk kedua model algoritma baik itu *cystitis* ataupun *pyelonephritis* bernilai $\geq 90\%$. Namun, terdapat kekurangan yang ditemukan saat proses pengujian sistem yaitu sistem tidak dapat melakukan mitigasi ketika terjadi *data imbalance*. Sebagai saran, sistem dapat diperbaiki dengan menambahkan proses *data preprocess* untuk mengatasi *data imbalance* sebelum training seperti *oversampling* atau *undersampling*. Saran lain adalah menambahkan *asynchronous process* untuk k-fold sehingga proses pelatihan pohon keputusan dapat dipercepat jika data yang ditambahkan sudah banyak. Data latih juga dapat menggunakan hasil test lab dan dikombinasikan dengan *dataset* yang ada. *Dataset* yang sudah ada bisa saja tidak sesuai dengan data lab yang baru nantinya, maka

dapat dilakukan pendekatan *majority vote* dari beberapa model kecerdasan buatan yang dibentuk oleh sistem.

DAFTAR PUSTAKA

- [1] C. Tan, and M. Chlebicki, "Urinary tract infections in adults", *Singapore Medical Journal*, Vol. 57, No. 9, pp. 485–490. 2016.
- [2] P.V. Darsono, D. Mahdiyah, F. Fahrianti, "Gambaran Karakteristik Ibu Hamil Yang Mengalami Infeksi Saluran Kemih (Isk) di Wilayah Kerja Puskesmas Pekauman Banjarmasin", *Dinamika Kesehatan*, Vol. 7, No. 1, pp. 162-170. 2016.
- [3] R. Li, and S.W. Leslie, "Cystitis", *Treasure Island (FL): StatPearls Publishing*. [Online], Available: <https://www.ncbi.nlm.nih.gov/books/NBK482435/>. [Accessed: Jul 3, 2021].
- [4] R. Li, and S.W. Leslie, "Acute Pyelonephritis", *Treasure Island (FL): StatPearls Publishing*. [Online], Available: <https://www.ncbi.nlm.nih.gov/books/NBK519537/>, [Accessed: Jul 3, 2021].
- [5] H. Llewelyn, H.A. Ang, K. Lewis, and A. Al-Abdullah, *Oxford handbook of clinical diagnosis*. New York, USA: Oxford University Press, 2014.
- [6] Z.H. Zhou, "A brief introduction to weakly supervised learning", *National*

- science review*, Vol. 5, No. 1, pp. 44-53. 2018.
- [7] C. Chazar, and B. Erawan, "Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine", *Jurnal Informatika dan Sistem Informasi*, Vol. 12, No. 1, pp. 67-80, 2020.
- [8] Z.A. Leleury, and B.P. Tomasouw, "Diagnosa Penyakit Saluran Pernapasan Dengan Menggunakan Support Vector Machine (SVM)", *Barekeng: Jurnal Ilmu Matematika dan Terapan*, Vol. 9, No. 2, pp 109-119. 2015
- [9] H.T. Sihotang, "Perancangan Aplikasi Sistem Pakar Diagnosa Diabetes Dengan Metode Bayes", *Jurnal Mantik Penusa*, Vol. 1, No. 1, 2017.
- [10] Y. Yuliyana, and A.S.R.M. Sinaga, "Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Naive Bayes". *Fountain of Informatics Journal*, Vol. 4, No. 1, pp. 19-23. 2019.
- [11] A. Mujahidin, and D. Pribadi, "Penerapan Algoritma C4. 5 untuk Diagnosa Penyakit Pneumonia Pada Anak Balita Berbasis Mobile". *Jurnal Swabumi*, Vol. 5, No. 2, pp. 155-161. 2017.
- [12] I.M. Nasser, and S.S. Abu-Naser, "Lung cancer detection using artificial neural network", *International Journal of Engineering and Information Systems*, Vol. 3, No. 3, pp. 17-23. 2019
- [13] J. Czerniak, and H. Zarzycki, "Application of rough sets in the presumptive diagnosis of urinary system diseases", In Proc. Artificial Intelligence and Security in Computing Systems, ACS'2002 9th International Conference Proceedings. 2003. pp. 41-51, 2003
- [14] J.R. Quinlan, *C45: Program for Machine Learning*, Massachusetts: Morgan Kaufmann Publisher, 1992.
- [15] W. Wiharto, H. Kusnanto, and Herianto, "Interpretation of Clinical Data Based on C4.5 Algorithm for the Diagnosis of Coronary Heart Disease". *Healthcare Informatics Research*, Vol. 22, No. 3, Jul, pp. 186-195. 2016.
- [16] V.S. Sowmien, K. Palani, T.R. Vijayaram, and V. Sugumaran, "Diagnosis of Hepatitis using Decision Tree Algorithm", *International Journal of Engineering and Technology*, Vol. 8, No. 3, pp. 1415-1419. 2016.
- [17] Sucipto, Kusrini, and E.L. Taufiq, "Classification method of multi-class on C4.5 algorithm for fish diseases", In 2016 2nd International Conference on Science in Information Technology (ICSITech), 2016, pp. 5-9.
- [18] K.A. Scorza, A. Williams, J.D. Phillips, and J. Shaw, "Evaluation of nausea and vomiting". *American family physician*, Vol. 76, No. 1, pp. 76-84. 2007.
- [19] W. Hochreiter, "Das schmerzhaftes Wasserlösen (Dysurie, Algurie) [Painful micturition (dysuria, alguria)]",

- Therapeutische Umschau. Revue therapeutique*, Vol. 53, No. 9, pp. 668–671. 1996.
- [20] T.C. Michels, and J. E. Sands, “Dysuria: evaluation and differential diagnosis in adults”, *American Family Physician*, Vol. 92, No. 9, pp. 778-786. 2015.
- [21] R. Parikh, A. Mathai, S. Parikh, G.C. Sekhar, and R. Thomas, “Understanding and using sensitivity, specificity and predictive values”, *Indian journal of ophthalmology*, Vol. 56, No. 1, pp. 45–50, 2008.
- [22] P. Chaurasia, “Associative Arrays in PHP”, 2021 [Online], Available: <https://www.geeksforgeeks.org/associative-arrays-in-php/>. [Accessed: Feb. 19, 2022].

Tabel Daftar istilah/Glosarium

No	Istilah	Penjelasan
1.	<i>Array associative</i>	Tipe data abstrak yang menyimpan kumpulan dari pasangan kunci atau index dan nilai. Berbeda dengan array biasa yang menggunakan nomer index posisi pada array untuk melakukan akses nilai array, <i>array associative</i> menggunakan key (biasanya merupakan tipe data <i>string</i>) untuk melakukan akses nilai.
2.	<i>Cystitis</i>	Infeksi saluran kemih bagian bawah, atau infeksi kantung kemih [3].
3.	Nausea	Rasa tidak nyaman, pusing, terkadang sampai mual dan sensasi sebelum muntah [18]
4.	<i>Mictur Pain</i>	Merupakan sensasi perih saat melakukan buang air kecil. Dapat diakibatkan oleh adanya <i>cystitis</i> [19,20]
5.	<i>Renal Pelvis</i>	Sebuah corong pada ginjal yang berfungsi untuk mengumpulkan air seni.
6.	<i>Pyelonephritis</i>	Infeksi bakteri yang menyebabkan peradangan pada ginjal. <i>Pyelonephritis</i> terjadi sebagai komplikasi dari infeksi saluran kemih bagian atas yang menyebar dari kandung kemih ke ginjal [4].