

IMPLEMENTASI *GREEDY FORWARD SELECTION* UNTUK PREDIKSI METODE PENYAKIT KUTIL MENGGUNAKAN *DECISION TREE*

Fitriyani¹, Toni Arifin²

¹Sistem Informasi, Universitas ARS, Bandung, Indonesia

²Teknik Informatika, Universitas ARS, Bandung, Indonesia

e-mail: fitriyani@ars.ac.id, toni.arifin@ars.ac.id

Abstrak

Penyakit kutil dapat ditangani dengan berbagai metode seperti *cryotherapy* dan *immunotherapy*, akan tetapi dokter belum mengetahui metode pengobatan yang paling tepat untuk pasien, sehingga diperlukan pengujian agar dapat diketahui metode yang paling tepat untuk pasien. Penelitian ini menggunakan dataset *cryotherapy* dan *immunotherapy* dengan menggunakan algoritma klasifikasi *Decision Tree*. Pada dataset ini terdapat atribut atau fitur yang tidak relevan sehingga dilakukan seleksi fitur menggunakan *Greedy Forward Selection*. Hasil penelitian ini akan dilakukan perbandingan kinerja dari algoritma *Decision Tree* tanpa seleksi fitur *Greedy Forward Selection* dengan *Decision Tree* yang diintegrasikan pada seleksi fitur *Greedy Forward Selection* dan pemilihan metode pengobatan penyakit kutil yang terbaik.

Kata kunci: Penyakit Kutil, *Cryotherapy*, *Immunotherapy*, *Decision Tree*, *Greedy Forward Selection*.

Abstract

Wart can be treated with various methods such as cryotherapy and immunotherapy, but doctors do not yet know the most appropriate treatment method for patients, so testing is needed in order to find out the most appropriate method for patients. This study uses a dataset of cryotherapy and immunotherapy using the Decision Tree classification algorithm. In this dataset there are irrelevant attributes or features, so a feature selection is done using Greedy Forward Selection. The results of this study will compare the performance of the Decision Tree algorithm without the Greedy Forward Selection feature selection with the Decision Tree which is integrated in the Greedy Forward Selection feature selection and the selection of the best treatment method for wart disease.

Keywords: *Warts, Cryotherapy, Immunotherapy, Decision Tree, Greedy Forward Selection.*

PENDAHULUAN

Metode pengobatan kutil yang biasanya dilakukan oleh dokter bisa bermacam-macam, mulai dari pemberian krim kulit dengan kandungan zat yang lebih kuat, *cryotherapy* atau pembekuan area kulit yang sakit dengan nitrogen, hingga menggunakan terapi laser. Dari beberapa metode yang dapat mengobati kutil ini, tidak satupun dapat menyembuhkan semua pasien, sehingga dokter harus mencari perawatan yang efektif dan dapat disesuaikan untuk setiap pasien

(Khozeimeh et al., 2017). Salah satu metode untuk mengobati kutil adalah metode *cryotherapy*, metode ini merupakan metode terbaru untuk menghilangkan virus dengan suhu ultra dingin yaitu (-196°C) (Astarini, Chappell, Scheuring, Thompson, & Miller, 2016). Sedangkan metode *immunotherapy* adalah pengobatan dengan meningkatkan sistem kekebalan tubuh untuk mengatasi penyakit kutil. *Immunotherapy* intralesi adalah pengobatan efektif untuk kutil.

Metode ini memiliki respons terapeutik yang lebih baik, membutuhkan lebih sedikit sesi, dan mampu mengobati kutil yang sudah lama (Khozeimeh et al., 2017).

Penelitian ini menggunakan dataset publik dari UCI Repository dengan nama dataset cryotherapy dan immunotherapy. Penelitian menggunakan dataset publik sangat dianjurkan karena sebanyak 64.79% penelitian menggunakan dataset publik dan sebanyak 35,21% penelitian menggunakan dataset privat (Wahono, 2015). Selain itu, penggunaan dataset publik dapat membuat penelitian berulang, terbantahkan dan diverifikasi (Catal, Diri, 2009).

Berbagai penelitian tentang pengobatan penyakit kutil memiliki permasalahan yaitu memilih metode terbaik untuk penanganannya, sehingga penelitian ini bertujuan untuk memilih metode terbaik dengan melakukan prediksi menggunakan algoritma *Decision Tree* dan *Greedy Forward Selection*. Algoritma *Decision Tree* termasuk dalam 10 algoritma terbaik (Wu & Kumar, 2009). Algoritma ini sangat komprehensif (mudah dipahami dan diinterpretasikan) ketika dibandingkan dengan yang lain karena algoritma pohon ini terdapat aturan yang dimana mudah dipahami dan diinterpretasikan (Gayatri, Nickolas, Reddy, 2010). Seleksi fitur dengan *Greedy Forward Selection* sangat efisien, sederhana dan tidak seperti teknik seleksi fitur yang membutuhkan waktu lama dalam prosesnya (*forward selection* dan

backward elimination). *Greedy Forward Selection* hanya memilih fitur yang berkontribusi dan dapat meningkatkan performa klasifikasi (Laradji, Alshayeb, Ghouti, 2015).

Banyak penelitian yang dilakukan sebelumnya dengan menggunakan dataset yang sama akan tetapi menggunakan metode yang berbeda, seperti yang dilakukan oleh (Nugroho, Adji, Setiawan, 2018), pada penelitian ini menggunakan metode *Random Forest Weighting based Feature Selection* dan C4.5. Penelitian lain menggunakan dataset yang berbeda, akan tetapi menggunakan algoritma yang sama C4.5 ada pada penelitian (Adeyemo & Adeyeye, 2015; Durairaj & Kalaiselvi, 2015; Rohman, Suhartono, & Supriyanto, 2017).

METODE

A. Dataset

Dataset yang digunakan dalam penelitian ini ada dua yaitu: *dataset cryotherapy* dan *immunotherapy*, kedua *dataset* ini dapat di akses pada situs UCI *Repository*.

Dataset cryotherapy dan *immunotherapy* merupakan *dataset* yang berisi tentang pengobatan penyakit kutil pada 90 pasien. *Dataset cryotherapy* berisi 7 atribut, sedangkan *dataset immunotherapy* berisi 8 atribut. Tabel 1 menunjukkan deskripsi atribut pada *dataset cryotherapy*, sedangkan Tabel 2 menunjukkan deskripsi atribut pada *dataset immunotherapy*.

Tabel 1. Deskripsi Atribut Cryotherapy

No.	Fitur atau Atribut	Dekripsi
1.	<i>Sex</i>	Jenis Kelamin Pasien (47 Laki-Laki, 43 Wanita)
2.	<i>Age</i>	Usia Pasien (15-67)
3.	<i>Time</i>	Waktu berlalu sebelum perawatan (<i>month</i> , 0-12)
4.	<i>Number_of_Warts</i>	Angka Kutil (1-12)
5.	<i>Type</i>	Jenis Kutil (1-Umum, 2-Plantar, 3-Keduanya)
6.	<i>Area</i>	Luas Permukaan Kutil (4-750)
7.	<i>Result_of_Treatment</i>	Hasil Pengobatan (Yes/No)

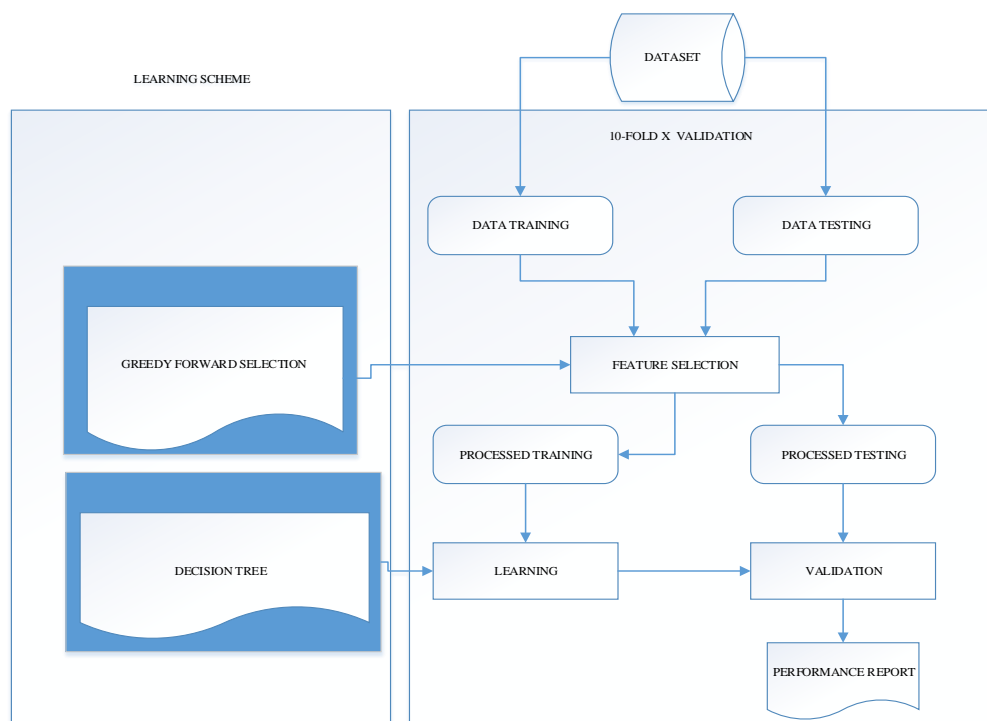
Tabel 2. Deskripsi Atribut *Dataset Immunotherapy*

No.	Fitur atau Atribut	Dekripsi
1.	<i>Sex</i>	Jenis Kelamin Pasien (41 Laki-Laki, 49 Wanita)
2.	<i>Age</i>	Usia Pasien (15-56)
3.	<i>Time</i>	Waktu berlalu sebelum perawatan (month, 0-12)
4.	<i>Number_of_Warts</i>	Angka Kutil (1-19)
5.	<i>Type</i>	Jenis Kutil (1-Umum, 2-Plantar, 3-Keduanya)
6.	<i>Area</i>	Luas Permukaan Kutil (6-900)
7.	<i>Induration_Diameter</i>	Diameter Indurasi Tes Awal (5-70)
8.	<i>Result_of_Treatment</i>	Hasil Pengobatan (Yes/No)

B. Model Penelitian

Model penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1. Pada Gambar 1 *Dataset* dibagi menjadi dua bagian, yaitu *data training* dan *data testing*, pembagian data ini dibagi secara otomatis menggunakan *10 fold cross validation*. *10 fold cross validation* merupakan pembagian *data training* dan *testing* menjadi 10 bagian

dan 10 validasi. Setelah data di validasi selanjutnya data akan diseleksi menggunakan seleksi fitur *Greedy Forward Selection* sehingga akan mendapatkan fitur atau atribut yang relevan. Kemudian fitur dan record akan diproses menggunakan algoritma *Decision Tree* untuk mendapatkan hasil dari kinerja model yang dibuat.



Gambar 1. Model Penelitian

1. *Feature Selection* (seleksi fitur)

Feature selection atau seleksi fitur dapat mengurangi dimensional pada data dan untuk meningkatkan kinerja dari mesin pembelajaran. *Subset selection* merupakan metode *feature selection*

(seleksi fitur), *subset selection* adalah menemukan subset terbaik (fitur), *subset* yang terbaik mempunyai jumlah dimensi yang paling berkontribusi pada akurasi. Metode ini akan menghilangkan sisanya atau dimensi yang tidak penting

menggunakan *error function* dan ini dapat menyelesaikan permasalahan pada regresi dan klasifikasi. Pemecahan masalah tersebut dapat diselesaikan dengan menggunakan metode *heuristic* untuk hasil yang masuk akal tapi bukan merupakan solusi yang optimal tapi masuk akal dalam waktu (Alpaydin Ethem, 2010).

Attribute subset selection adalah mengurangi ukuran *dataset* dengan menghilangkan atribut yang tidak relevan atau perulangan atribut (dimensi). Dalam *machine learning*, *Attribute subset selection* dapat disebut *feature subset selection*. Tujuan *attribute subset selection* adalah untuk menemukan satu set minimal atribut sehingga distribusi probabilitas yang dihasilkan dari kelas data sedekat mungkin dengan distribusi asli yang diperoleh dengan menggunakan semua atribut (Han, Kamber, & Pei, 2012).

Ada banyak metode yang digunakan dalam seleksi atribut atau fitur seperti: *Greedy Forward Selection*, *greedy backward elimination*, *forward selection*, *backward elimination*, *information gain* dan lain-lain. Seleksi fitur dengan *forward selection* adalah dimulai dengan tidak adanya atribut kemudian pada setiap langkah ditambahkan satu persatu yang memiliki kontribusi dan tidak menggunakan atribut yang tidak berkontribusi. Seleksi fitur dengan *backward elimination* adalah dengan seleksi mundur, maksudnya adalah semua atribut digunakan dan dihapus satu persatu hingga mendapatkan atribut yang menghasilkan kontribusi.

2. Algoritma Greedy

Algoritma *greedy* merupakan algoritma yang menggunakan *metaheuristic* untuk menyelesaikan permasalahan dengan mengidentifikasi menggunakan lokal optimum dan mengarah pada global optimum dalam pendekatannya (Gorunescu, 2011).

Algoritma *greedy* merupakan metode yang paling populer untuk memecahkan persoalan optimasi (*optimization problems*) atau dapat disebut mencari solusi optimum. Algoritma *greedy* dapat diartikan sebagai rakus, tamak, loba dan lain-lain dengan menggunakan prinsip *take what you can get now*, algoritma ini membentuk solusi langkah per langkah (*step by step*).

Pada setiap langkahnya terdapat banyak pilihan yang perlu dieksplorasi dan dari setiap langkah tersebut harus dibuat keputusan yang terbaik dalam menentukan pilihan. Dalam setiap langkah, algoritma ini membuat pilihan optimum lokal untuk dapat mengarah pada solusi optimum global.

Contoh pemecahan masalah penukaran uang menggunakan algoritma *greedy*:

Koin: 2, 3, 1, 4

Uang yang ditukar: 10

Solusi *greedy*: pada setiap langkah, memilih koin dengan nilai terbesar dari himpunan koin

Langkah 1: 2+4+4 (total 3 koin)

Langkah 2: 2+2+3+3 (total 4 koin)

Langkah 3: 2+1+1+4+2 (total 5 koin)

Langkah 3 merupakan solusi yang digunakan oleh *greedy*.

3. Greedy Forward Selection

Greedy Forward Selection merupakan salah satu pendekatan dalam algoritma *greedy* pada metode seleksi atribut atau seleksi fitur. Metode seleksi atribut *greedy* terdapat beberapa pendekatan seperti: *greedy forward selection*, *greedy backward elimination* dan kombinasi *forward selection* dan *backward elimination* (Han et al., 2012). Algoritma *greedy* dapat disebut juga dengan *stepwise* dan weka menggunakan algoritma ini dengan sebutan *stepwise*. Pada Gambar 2 dijelaskan mengenai metode seleksi fitur dengan *greedy* dimana untuk mencari atribut terbaik atau terburuk dapat menggunakan pengukuran *information gain* (Han et al., 2012).

Algoritma *greedy* dengan seleksi *subset atribut* (Han et al., 2012) sebagai berikut untuk penjelasannya:

a. Stepwise forward selection

Prosedur dimulai dengan himpunan kosong dari atribut sebagai set yang dikurangi, atribut yang terbaik dari atribut asli ditentukan dan ditambahkan pada set yang kurang. Pada setiap iterasi berikutnya yang terbaik dari atribut asli yang tersisa ditambahkan ke set.

b. Stepwise backward elimination

Prosedur dimulai dengan full set atribut. Pada setiap langkah, teknik ini

dapat menghilangkan atribut terburuk yang tersisa di *dataset*.

- c. Kombinasi *forward selection* dan *backward elimination*

Stepwise forward selection dan *backward elimination* dapat dikombinasikan. Pada setiap langkah, prosedur memilih atribut terbaik dan menghilangkan yang terburuk dari atribut yang tersisa.

4. *Decision Tree*

Decision Tree merupakan serangkaian pertanyaan yang sistematis dan semuanya diatur sehingga setiap pertanyaan dari *query* atribut (misalnya *Outlook*) berdasarkan nilai dari atribut (Wu & Kumar, 2009). Algoritma C4.5 dirancang oleh J. Ross Quinlan, dinamakan demikian karena merupakan keturunan dari pendekatan ID3 untuk merangsang pohon keputusan.

Algoritma C4.5

Input: Dataset atribut bernilai D

- 1: $Tree = \{\}$
- 2: **if** D adalah "murni" OR kriteria lainnya akan terpenuhi **then**
- 3: terminate
- 4: **end if**
- 5: **for all** atribut $a \in D$ **do**
- 6: menghitung kriteria informasi-teori jika kita membagi pada a
- 7: **end for**
- 8: a_{best} = atribut terbaik sesuai dengan kriteria yang dihitung
- 9: $Tree$ = buat simpul keputusan yang menguji a_{best} di root
- 10: D_v = induksi sub dataset dari D berbasis a_{best}
- 11: **for all** D_v **do**
- 12: $Tree_v = C4.5(D_v)$
- 13: melampirkan $Tree_v$ untuk cabang pohon yang sesuai
- 14: **end for**
- 15: **return** $Tree$

Pada *Decision Tree* untuk menentukan akar dari pohon adalah

dengan mengambil atribut yang terpilih dengan cara menghitung gain dari masing-masing atribut. Nilai gain paling tertinggi menjadi akar pertama. Sebelum menghitung gain, harus menghitung entropy dengan rumus sebagai berikut:

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

Keterangan:

S = Himpunan Kasus

n = Jumlah Partisi S

p_i = Proporsi S_i Terhadap S

Setelah mendapatkan nilai entropy dari masing-masing atribut, selanjutnya menghitung nilai gain dengan menggunakan rumus:

$$Gain(S, A) = Entropy(S)$$

$$- \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i)$$

Keterangan:

S = Himpunan Kasus

A = Fitur

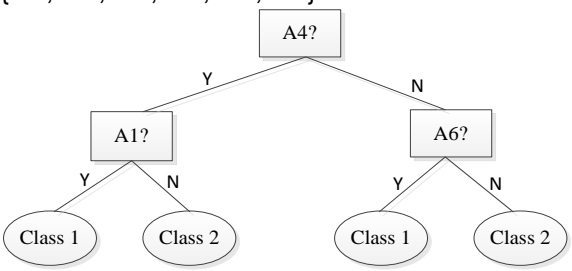
n = Jumlah Partisi Atribut A

$|S_i|$ = Proporsi S_i Terhadap S

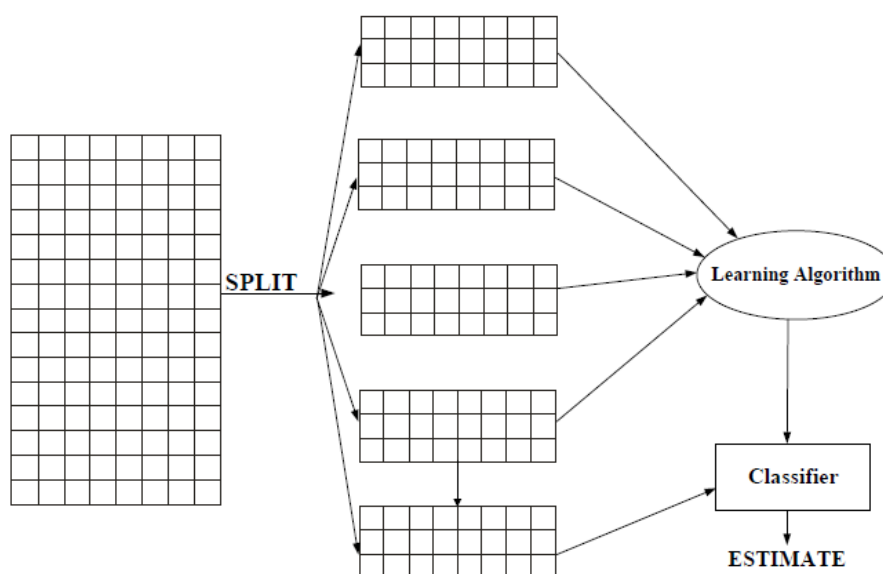
$|S|$ = Jumlah Kasus dalam S

5. *K-Fold Cross Validation*

Cross validation digunakan ketika tujuan dari machine learning adalah prediksi dan sangat penting untuk mengevaluasi kinerja metode *machine learning* (Sumeet & Xian, 2011). Dataset terdiri kasus N , ini dibagi menjadi bagian-bagian k sama, k biasanya menjadi sejumlah kecil seperti 5 atau 10. (Jika N tidak tepat habis dibagi oleh k , bagian akhir akan memiliki kasus lebih sedikit dari k lain - 1 bagian). Serangkaian berjalan k kini dilakukan. Setiap bagian k pada gilirannya digunakan sebagai ujian menetapkan dan k lainnya - 1 bagian digunakan sebagai training set (Bramer, 2016). Gambar 3 merupakan pembagian data training dan data testing pada dataset dengan menggunakan k -fold cross validation.

Forward Selection	Backward Elimination	Decision Tree Induction
Atribut asli: {A1, A2, A3, A4, A5, A6} Pengurangan atribut asli: {} =>{A1} =>{A1, A4} =>Pengurangan atribut: {A1, A4, A6}	Atribut asli: {A1, A2, A3, A4, A5, A6} =>{A1, A3, A4, A5, A6} =>{A1, A4, A5, A6} =>Pengurangan Atribut: {A1, A4, A6}	Atribut asli: {A1, A2, A3, A4, A5, A6}  =>Pengurangan Atribut: {A1, A3, A6}

Gambar 2. Metode *Heuristic (Greedy)* untuk Seleksi Subset Atribut
 Sumber: (Han et al., 2012)



Gambar 3. *K-Fold Cross Validation*
 Sumber: (Bramer, 2016)

HASIL DAN PEMBAHASAN

Pada penelitian ini eksperimen dilakukan sebanyak dua kali pada masing masing dataset, dimana eksperimen pertama dataset diuji dengan algoritma *Decision Tree* tanpa seleksi fitur *Greedy Forward Selection*, kemudian *dataset* diuji lagi dengan algoritma *Decision Tree* dan seleksi fitur *Forward Selection*. Kemudian hasil dari keduanya dibandingkan untuk mendapatkan nilai kinerja yang terbaik. Tabel 3 menunjukkan hasil kinerja dari model *Decision Tree* tanpa seleksi fitur *Greedy Forward Selection* menggunakan dataset *cryotherapy*. TF adalah True False,

TT adalah *True True*, sedangkan PF adalah *Prediction False* dan PT adalah *Prediction True*. Pada eksperimen ini hasil akurasi yang didapat adalah 90%. *Rule model Decision Tree* ditunjukkan pada Gambar 4.

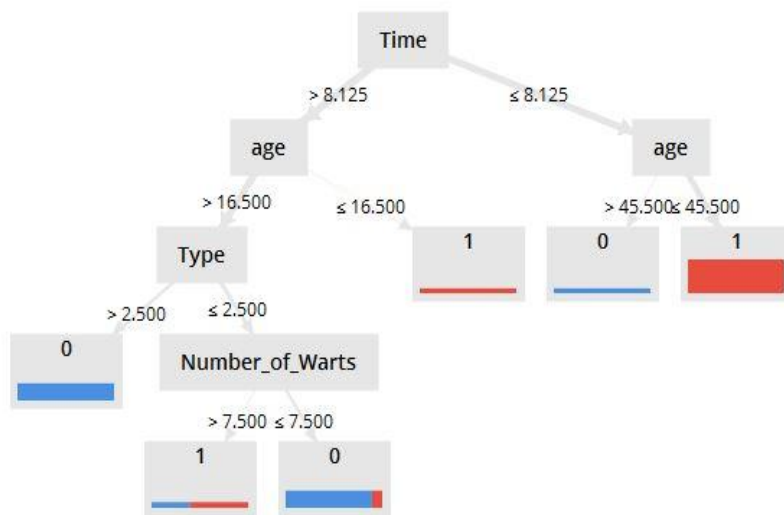
Tabel 4 menunjukkan hasil dari kinerja model yang menggunakan algoritma *Decision Tree* dan *Greedy Forward Selection*. Pada model ini hasil kinerja atau akurasi yang didapat adalah 92.22%. Pada eksperimen ini hasil seleksi fitur dengan *Greedy Forward Selection* adalah fitur *Age* dan *Time*.

Selanjutnya eksperimen pada *dataset Immunotherapy*, pada eksperimen pertama *dataset immunotherapy* menggunakan model *Decision Tree* tanpa seleksi fitur *Greedy Forward Selection*. Tabel 5 menunjukkan hasil kinerja dari

model *Decision Tree* pada *dataset Immunotherapy*. Hasil akurasi yang didapat pada eksperimen menggunakan model *Decision Tree* adalah 78.89%. Gambar 5 merupakan model *rule Decision Tree* pada *dataset Immunotherapy*.

Tabel 3. Hasil Kinerja *Decision Tree Dataset Cryotherapy*

	TF	TT	Class Precision
PF	38	5	88.37%
PT	4	43	91.49%
Class Recall	90.48%	89.58%	



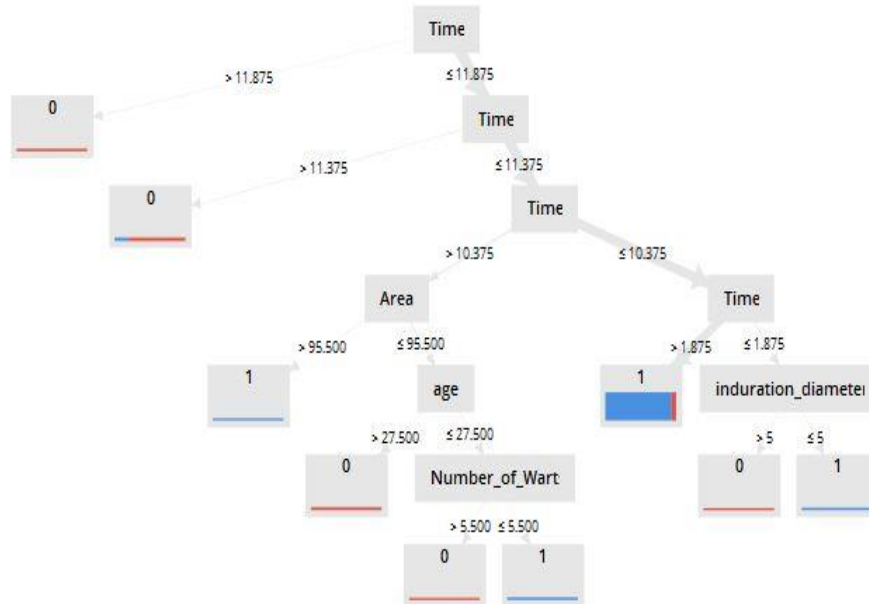
Gambar 4. Rule Model *Decision Tree Cryotherapy*

Tabel 4. Hasil Kinerja *Decision Tree dan Greedy Forward Selection Dataset Cryotherapy*

	TF	TT	Class Precision
PF	40	5	88.89%
PT	2	43	95.56%
Class Recall	95.24%	89.58%	

Tabel 5. Hasil Kinerja *Decision Tree Dataset Immunotherapy*

	TF	TT	Class Precision
PF	65	13	83.33%
PT	6	6	50.00%
Class Recall	91.55%	31.58%	



Gambar 5. Rule Model Decision Tree Dataset Immunotherapy

Pada eksperimen selanjutnya menggunakan model algoritma Decision Tree dan *Greedy Forward Selection*. Tabel 6 merupakan hasil kinerja model Decision Tree dan seleksi fitur *Greedy Forward Selection* dengan menggunakan dataset

Immunotherapy. Hasil akurasi menggunakan model ini adalah 82,22%. Seleksi fitur yang paling relevan menggunakan *Greedy Forward Selection* adalah fitur Time.

Tabel 6. Hasil Kinerja Decision Tree dan Greedy Forward Selection

	TF	TT	Class Precision
PF	65	10	86,67%
PT	6	9	60,00%
Class Recall	91,55%	47,37%	

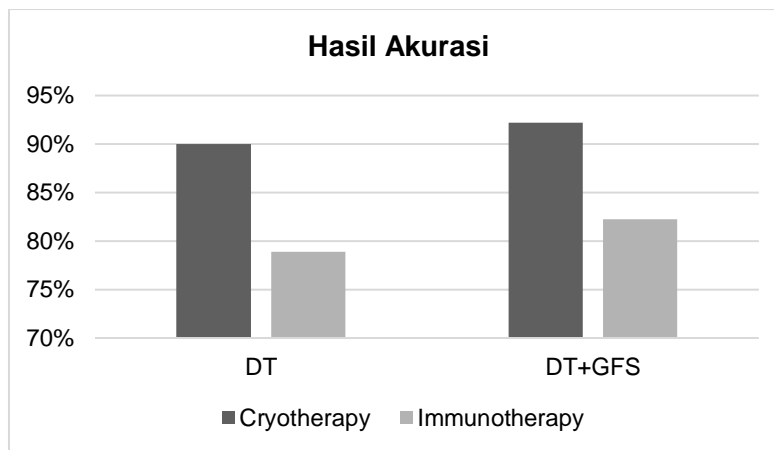
Tabel 7 menunjukkan perbandingan dari kedua model yang digunakan dalam penelitian ini pada dua dataset yang berbeda. Pada dataset cryotherapy nilai akurasi tertinggi adalah sebesar 92.22% menggunakan model Decision Tree (DT) dan seleksi fitur Greedy Forward Selection (GFS), sedangkan pada dataset immunotherapy nilai akurasi tertinggi

adalah 82,22% menggunakan model Decision Tree (DT) dan Greedy Forward Selection (GFS).

Gambar 6 menunjukkan perbandingan grafik nilai akurasi dari kedua model dan kedua dataset. Grafik ini menunjukkan bahwa seleksi fitur Greedy Forward Selection dapat meningkatkan hasil kinerja dari model yang digunakan.

Tabel 7. Perbandingan Nilai Akurasi

Dataset	DT	DT+GFS
Cryotherapy	90%	92.22%
Immunotherapy	78.89%	82.22%



Gambar 6. Perbandingan Model DT dan DT+GFS

Tabel 8 merupakan perbandingan hasil kinerja *Decision Tree* (DT) dan *Greedy Forward Selection* (GFS) dengan hasil penelitian yang lain. Pada penelitian Nugroho et al. (2018) dengan menggunakan algoritma C4.5+RFFW (*Random Forest Feature Weigthing*), hasil terbaik terdapat pada dataset cryotherapy, sedangkan penelitian Khozeimeh et al. (2017) menggunakan algoritma ANFIS (*Adaptive Network-based Fuzzy Inference System*), hasil terbaik terdapat pada dataset cryotherapy. Penelitian lain Basarslan & Kayaalp (2018) menggunakan algoritma *Correlation Based Feature Selection* (CFS) dan C4.5, hasil terbaik terdapat pada dataset cryotherapy.

Tabel 8. Perbandingan Hasil dengan Penelitian yang lain

Algoritma	Cryotherapy	Immunotherapy
DT	90%	78,89%
DT+GFS	92,22%	82,22%
C4.5+RFFW	93,33%	84,44%
ANFIS	83,33%	80%
CFS+C4.5	95,4%	84%

SIMPULAN

Penyakit kutil dapat ditangani dengan berbagai metode seperti *cryotherapy* dan *immunotherapy*, akan tetapi perlu dilakukan pengujian untuk mengetahui metode mana yang paling sesuai dengan pasien. *Dataset cryotherapy* dan *immunotherapy* memiliki fitur yang tidak relevan sehingga dibutuhkan algoritma seleksi fitur yang dapat meningkatkan kinerja dari model penelitian. Hasil penelitian pada dataset *cryotherapy* menggunakan *Decision Tree* mendapatkan nilai akurasi 90%, sedangkan hasil seleksi fitur menggunakan *Greedy Forward Selection* sebesar 92,22%. Hasil penelitian pada dataset *immunotherapy* menggunakan *Decision Tree* tanpa seleksi fitur mendapatkan nilai akurasi sebesar 78,89%, sedangkan menggunakan seleksi fitur *Greedy Forward*

Selection mendapat nilai akurasi sebesar 82,22%. Hasil ini menunjukkan bahwa seleksi fitur *Greedy Forward Selection* dapat meningkatkan kinerja dari model.

UCAPAN TERIMAKASIH

Peneliti mengucapkan terima kasih kepada RISTEKDIKTI atas dukungan dana penelitian pada hibah penelitian dosen pemula 2020 dan peneliti juga mengucapkan terimakasih kepada Fahime Khozeimeh, Pouran Layegh, Roohallah Alizadehsani, Mohamad Roshanzamir atas dataset *cryotherapy* dan *immunotherapy* di UCI Repository.

DAFTAR PUSTAKA

Adeyemo, O. ., & Adeyeye, T.2015. Comparative Study of ID3/C4. 5 Decision tree and Multilayer

- Perceptron Algorithms for the Prediction of Typhoid Fever. *African Journal of Computing & ICT*, 8(1), 103–112.
- Alpaydin Ethem. 2010. *Introduction to Machine Learning Second Edition* (2nd ed.; T. Dietterich, ed.). London: MIT.
- Astarini, I. A., Chappell, A., Scheuring, D., Thompson, S. M., & Miller, C. 2016. Optimasi Metode Cryotherapy untuk Mengeliminasi Virus pada Tunas Kentang In Vitro (Optimization of Cryotherapy Method to Eliminate Virus on In Vitro Potato Shoot Tips). *J. Hort*,
- Basarslan, M. S., & Kayaalp, F. 2018. A Hybrid Classification Example in the Diagnosis of Skin Disease with Cryotherapy and Immunotherapy Treatment. *ISMSIT 2018 - 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies, Proceedings*, 1–5.
- Bramer, M. 2016. *Principles of Data Mining*.
- Catal, C., & Diri, B. 2009. A systematic review of software fault prediction studies. *Expert Systems with Applications*, 36(4), 7346–7354.
- Durairaj, M., & Kalaiselvi, G. 2015. Prediction Of Diabetes Using Soft Computing Techniques-A Survey. *International Journal Of Scientific & Technology Research*, 4(03), 190–192.
- Gayatri, N., Nickolas, S., & Reddy, A. V. 2010. Feature Selection Using Decision Tree Induction in Class level Metrics Dataset for Software Defect Predictions. *Proceedings of the World Congress on Engineering and Computer Science 2010, I*.
- Gorunescu, F. 2011. Data mining: concepts and techniques. In *Springer*.
- Han, J., Kamber, M., & Pei, J. 2012. *Data Mining: Concepts and Techniques*.
- Khozeimeh, F., Alizadehsani, R., Roshanzamir, M., Khosravi, A., Layegh, P., & Nahavandi, S. 2017. An expert system for selecting wart treatment method. *Computers in Biology and Medicine*, 81, 167–175.
- Laradji, I. H., Alshayeb, M., & Ghouti, L. 2015. Software defect prediction using ensemble learning on selected features. *Information and Software Technology*, 58, 388–402.
- Nugroho, H. W., Adji, T. B., & Setiawan, N. A. 2018. *Random Forest Weighting based Feature Selection for C4 . 5 Algorithm on Wart Treatment Selection Method*. 8(5), 1858–1863.
- Rohman, A., Suhartono, V., & Supriyanto, C. 2017. Penerapan Algoritma C4.5 berbasis Adaboost untuk Prediksi Penyakit Jantung. *Jurnal Teknologi Informasi*, 13, 13–19.
- Sumeet, D., & Xian, D. 2011. *Data Mining and Machine Learning in Cybersecurity*.
- Wahono, R. S. 2015. A Systematic Literature Review of Software Defect Prediction: Research Trends, Datasets, Methods and Frameworks. *Journal of Software Engineering*, 1(1).
- Wu, X., & Kumar, V. 2009. *The Top Ten Algorithms in Data Mining* (V. Kumar, ed.). Boca Raton, London, New York: CRC Press taylor & Francis Group.