

PREDIKSI KEBERHASILAN IMMUNOTHERAPY PADA PENYAKIT KUTIL DENGAN MENGGUNAKAN ALGORITMA NAÏVE BAYES

Toni Arifin¹, Siti Syalwah²

¹Universitas Adhirajasa Reswara Sanjaya
e-mail: toni.arifin@ars.ac.id

²Universitas Adhirajasa Reswara Sanjaya
e-mail: sitisyalwah13@gmail.com

Abstrak

Penyakit kutil merupakan masalah kesehatan kulit yang pada umumnya ditandai dengan munculnya benjolan kecil bertekstur kasar pada permukaan kulit yang disebabkan oleh virus yaitu *Human Papilloma Virus* (HPV). Salah satu teknik pengobatan penyakit kutil adalah *Immunotherapy*, metode ini merupakan pengobatan dengan meningkatkan sistem kekebalan tubuh untuk mengatasi penyakit kutil. Untuk mengurangi kesalahan deteksi dan menghindari keterlambatan diagnosis penderita penyakit kutil dapat dilakukan penerapan dan pemanfaatan teknik *data mining*. Salah satu metode *data mining* yang bisa dimanfaatkan adalah klasifikasi dan metode algoritma yang dapat digunakan untuk klasifikasi adalah algoritma *Naive Bayes*. Pada penelitian ini, algoritma klasifikasi *data mining Naive Bayes* diterapkan pada *Immunotherapy Data Set*. Setelah hasil penelitian diperoleh, terbukti metode *data mining* menggunakan algoritma *Naive Bayes* dengan *10-Fold Cross Validation* menghasilkan nilai akurasi klasifikasi sebesar 81,11% dan nilai *Area Under ROC* (AUC) sebesar 0,63 yang termasuk kategori *Good Classification*.

Kata Kunci: Penyakit Kutil, *Immunotherapy*, *Data Mining*, Klasifikasi, *Naive Bayes*

Abstract

Warts Disease is a skin health problem that is generally characterized by the appearance of small, rough textured lumps on the skin surface caused by a virus that is Human Papilloma Virus (HPV). One technique of treatment of wart disease is Immunotherapy, this method is a treatment by boosting the immune system to overcome the disease of warts. To reduce detection errors and avoid delays in diagnosing Warts disease, the application of data mining techniques can be applied and used. One method of data mining that can be used is classification and algorithm methods that can be used for classification are the Naive Bayes algorithm. In this study, the Naive Bayes data mining classification algorithm was applied to Immunotherapy Data Set. After the research results were obtained, it has been proven that data mining method using Naive Bayes algorithm with 10-Fold Cross Validation resulted in classification accuracy values of 81,11% and Area Under ROC (AUC) values of 0.63 which belonged to the Good Classification category.

Keywords: *Warts Disease, Immunotherapy, Data Mining, Classification, Naive Bayes*

1. Pendahuluan

Penyakit kutil merupakan salah satu masalah kesehatan kulit yang biasanya ditandai dengan timbulnya benjolan kecil pada permukaan kulit, penyakit ini yang disebabkan oleh virus yaitu *Human Papiloma Virus* (HPV). Penularan virus penyebab kutil dapat terjadi dengan mudah, salah satunya hanya bersentuhan langsung dengan

seseorang penderita kutil, namun tidak semua orang yang bersentuhan dengan virus *Human Papiloma Virus* (HPV) akan menimbulkan penyakit kutil. Imunitas tubuh masing-masing orang sangat berpengaruh dalam penularan penyakit ini, penderita yang memiliki imunitas yang bermasalah akan rentan untuk terserang penyakit kutil. Salah satu teknik pengobatan kutil adalah

dengan *Immunotherapy*, metode ini merupakan pengobatan dengan meningkatkan sistem kekebalan tubuh untuk mengatasi penyakit kutil. *Immunotherapy intralesi* adalah pengobatan efektif untuk kutil. Metode ini memiliki respon terapeutik yang lebih baik membutuhkan lebih sedikit sesi, dan mampu mengobati kutil yang sudah lama (Supriyatna & Mustika, 2018).

Human Papilloma Virus (HPV) adalah virus yang paling sering dijumpai pada penyakit menular seksual dan diduga berperan dalam proses terjadinya kanker. Terdapat sekitar 130 tipe *Human Papilloma Virus* (HPV) yang telah diidentifikasi (Setiawati, 2014).

Immunotherapy pertama kali diperkenalkan oleh William B Colley. Colley mencatat bahwa beberapa pasien yang mengidap sarcoma ini mengalami regresi pada tumor secara spontan dan ini berkorelasi dengan infeksi bakteri. Colley kemudian menggunakan bakteri untuk menginfeksi pasien kanker dan dalam beberapa kasus, baru-baru ini peran sistem kekebalan tubuh dalam penyakit kanker lebih didefinisikan sebagai *Immunoediting*

Immunotherapy juga telah dievaluasi dalam gabungan dengan terapi lain seperti pembedahan, kemoterapi, dan terapi radiasi. Bila digunakan dalam gabungan dengan modalitas sitoreduktif lain ini, maka terapi dinamai sebagai *Immunotherapy* (Ramadhan & Mahmudy, 2019).

Immunotherapy adalah salah satu penanganan pada penyakit yang disebabkan oleh virus *Human Papilloma Virus* (HPV). Metode ini ditemukan pada tahun 1970-an untuk penyembuhan kanker melanoma kandung kemih dengan menggunakan agen imun *Cytokine* namun memiliki efek samping yang besar dan seiring mempelajari respon imun, metode ini terus dikembangkan. Metode *Immunotherapy* merangsang beberapa sistem kekebalan tubuh manusia dan penggunaan metode ini terbukti ampuh untuk penanganan penyakit yang disebabkan oleh *Human Papilloma Virus* (HPV) (Ramadhan & Mahmudy, 2019).

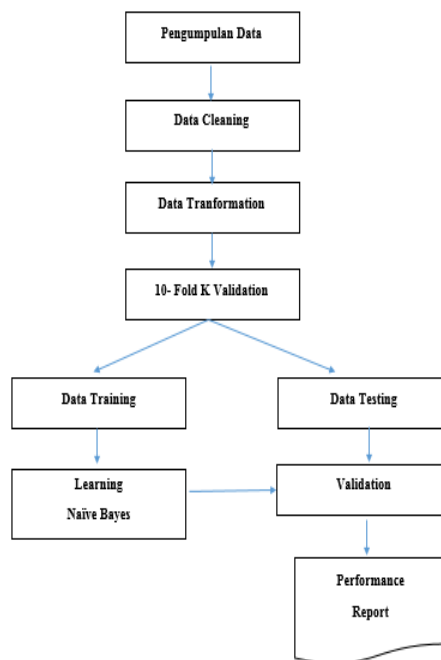
Penelitian terkait digunakan sebagai acuan dan referensi yang dapat memudahkan peneliti dalam menentukan langkah sistematis dari berbagai teori konseptual yang akan dikaji. Penelitian terkait berfungsi untuk menghindari duplikasi dan penanggulangan penelitian atau kesalahan yang sama seperti yang dibuat

oleh peneliti sebelumnya (Hermawan, Sukma, & Halfis, 2019) Berikut penelitian yang juga pernah membahas tentang prediksi keberhasilan *Immunotherapy* pada penyakit kutil :

Penelitian Hermawan, Sukma, Halfis Penelitian yang berjudul Analisis Algoritma Klasifikasi C.45 Untuk Memprediksi Keberhasilan *Immunotherapy* Pada Penyakit Kutil. Dimana dari hasil pengujian menggunakan algoritma naïve bayes dapat memprediksi *class* untuk semua *instance* secara tepat dengan nilai akurasi 81,11%.

2. Metode Penelitian

Pada penelitian ini diperlukan kerangka penelitian yang akan dilakukan, dibawah ini adalah gambar 1 Kerangka penelitian



Gambar 1. Kerangka Penelitian

- Pengumpulan Data**
Pada penelitian ini penulis mendapatkan data dari *Website University of California Irvine Machine learning Data Repository (UCI Repository)* pada tahun 2018. Dataset ini terdiri dari 90 Record dengan 8 atribut.
- Data Cleaning**
Pada *Immunotherapy Data Set*, penulisan pada class dari diagnosis menggunakan tipe numerikal 1 (*Berhasil*) dan 0 (*Gagal*). Algoritma

Naïve Bayes tidak bisa mengkategorikan *dataset* dengan *class* bertipe numerikal. Sehingga dilakukan *data cleaning* dengan mengubah numerikal 1 menjadi teks Berhasil, dan numerikal 0 menjadi teks Gagal.

c. Data Transformasi

Pada penelitian ini menggunakan *Immunotherapy Data Set* dengan jumlah 90 *record* dan 8 atribut. Namun, untuk proses *data mining* menggunakan algoritma *Naive Bayes* akan lebih optimal bila semua variabel atau atribut tidak mengandung nilai *continuous* dan bernilai kategorikal. Sehingga diperlukan proses transformasi data pada tahap pengolahan data awal. Pada *immunotherapy Disease Data Set* terdiri dari variabel campuran antara variabel diskrit atau kategorikal dan variabel *continuous* atau numerikal. Proses transformasi yang akan digunakan adalah *Discretization*. Dalam tahap *Discretization*, nilai-nilai baku dari atribut numerik (misalnya, *age*) akan diganti dengan label *Interval* (10-20 , 21-30). Proses ini dilakukan untuk mengubah semua variabel baku atau *continuous* menjadi variabel dengan nilai interval atau kategorikal.

Di bawah ini adalah tabel yang menjelaskan jumlah data yang dipakai dan nama atribut yang ada pada *Immunotherapy Data Set*.

Tabel 1. Nama-nama atribut

No	Atribut
1	Sex
2	Age
3	Time
4	Number of Warts
5	Type
6	Area
7	Induration Diameter
8	Result of Treatment

Tabel 2. jumlah data 2 kelas dataset

Kelas Berhasil	71
Kelas Gagal	19
Total	90

d. Langkah selanjutnya pembagian data set *Immunotherapy Dataset* menjadi 10 *Fold cross validation*,

semua data set akan dibagi menjadi data *training* dan data *testing*

- e. Langkah selanjutnya setelah didapatkan atribut terbaik adalah tahapan klasifikasi data dengan metode klasifikasi data mining
- f. Setelah data *training* dan data *testing* divalidasi dengan menggunakan metode klasifikasi. Langkah terakhir yaitu membandingkan hasil klasifikasi untuk mengetahui metode terbaik untuk klasifikasi dataset *Immunotherapy dataset*

Naive Bayes Classifier

Naive Bayes Classifier adalah salah satu algoritma dalam teknik *data mining* yang menerapkan *Bayesian Theorem* dalam klasifikasi. *Naive Bayes Classifier* merupakan pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class*. *Naive Bayes Classifier* terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* dengan data yang besar lainnya (Herliana, Arifin, Susanti, & Hikmah, 2018). Persamaan *Teorema Bayes* (Arifin, 2015)

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Keterangan :

B : Data dengan *class* yang belum diketahui
A : Hipotesis data merupakan suatu *class* spesifik

P(A|B) : Probabilitas hipotesis A berdasarkan kondisi B (*posterior probabilities*)

P(A) : Probabilitas hipotesis A (*prior probabilities*)

P(B|A) : Probabilitas B Berdasarkan hipotesis A (*Conditional Probabilities*)

P(B) : Probabilitas B

Rumus diatas disesuaikan untuk *Naive Bayes* sebagai berikut:

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)}$$

$P(B)$ harus dimaksimumkan karena nilainya sama untuk semua kelas, sehingga rumus diatas disederhanakan lagi ke dalam bentuk seperti dibawah ini:

$$P(A_i|B) = P(B|A_i) \cdot P(A_i)$$

Asumsi sederhana dalam *Naive Bayes* adalah semua atribut kondisinya independen. Jadi tahapan rumusan kelas bagi data tes adalah berdasarkan rumus berikut:

$$P(B|Ai) = \prod_{k=1}^n P(Bk|Ai)$$

$$\text{Arg max}_{Ai} \{P(B|Ai)P(Ai)\}$$

Sebagai contoh bila ada data baru dan nilai posterior probabilitasnya $P(A_2|B)$ memiliki nilai paling besar di antara semua posterior probabilitas $P(A_k|B)$ untuk sejumlah k kelas, maka data tersebut termasuk ke dalam kelas A_2 .

Evaluasi dan Validasi

Validasi adalah proses mengevaluasi akurasi dari sebuah model, validasi mengacu untuk mendapatkan prediksi dengan menggunakan model yang ada kemudian membandingkan hasil yang diperoleh dengan hasil yang diketahui (Gorunescu, 2011). Mengevaluasi akurasi dari model klasifikasi sangat penting, akurasi dari sebuah model mengindikasikan kemampuan model tersebut untuk memprediksi *class target* (Vercellis, 2009). Untuk mengevaluasi model digunakan metode *Confusion Matrix* dan Kurva ROC (*Receiver Operating Characteristic*) (Arifin, 2015)

a. Confusion Matrix

Evaluasi kinerja model klasifikasi didasarkan pada pengujian objek yang diprediksi dengan benar dan salah, hitungan ini ditabulasikan *Confusion Matrix* (Gorunescu, 2011). *Confusion Matrix* adalah alat yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda (Han, Micheline, & Jian, 2012)

b. Kurva ROC

Kurva ROC banyak digunakan untuk menilai hasil prediksi, kurva ROC adalah teknik untuk memvisualisasikan, mengatur, dan memilih pengklasifikasian berdasarkan kinerja algoritma. Untuk klasifikasi *data mining*, nilai *Area Under ROC* (AUC) dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011):

1. 0.90 – 1.00 = *Excellent Classification*
2. 0.80 – 0.90 = *Good Classification*
3. 0.70 – 0.80 = *Fair Classification*
4. 0.60 – 0.70 = *Poor Classification*
5. 0.50 – 0.60 = *Failure*

3. Hasil dan Pembahasan

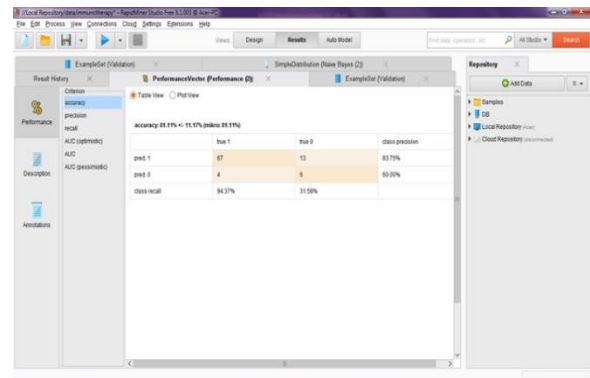
Penerapan algoritma *Naive Bayes* pada *Immunotherapy Data Set* dimaksudkan untuk mengetahui dan mendapatkan hasil akurasi lebih baik pada klasifikasi deteksi penyakit Kutil dari penelitian-penelitian sebelumnya. Eksperimen pada algoritma *Naive Bayes* menggunakan algoritma *Naive Bayes* dengan menggunakan metode validasi *K-Fold Cross Validation*.

a. Hasil Eksperimen *Naive Bayes* dengan *K-Fold Cross Validation*

Tabel 3. Hasil Akurasi Model *Naive Bayes* Menggunakan *K-Fold Cross Validation*

Model (Algoritma)	Jumlah <i>Fold Cross Validation</i>	Akurasi
<i>Naive Bayes</i>	5	80,00 %
<i>Naive Bayes</i>	10	81,11 %

Berikut merupakan hasil evaluasi terdapat pada gambar 2.



Gambar 2. Hasil Evaluasi Model *K-Fold Validation* dengan 10 *K-fold Validation*

Dari *confusion matrix* pada Gambar IV.1 dapat diukur tingkat akurasi dari klasifikasi sebagai berikut:

$$TP = 67; FP = 4; TN = 6; FN = 13$$

$$\text{Akurasi} = \frac{(TP+TN)}{(TP+TN+FP+FN)} * 100$$

$$\text{Akurasi} = \frac{(67+6)}{(67+6+4+13)} * 100$$

$$\text{Akurasi} = 81,11\%$$



Gambar 3. Kurva ROC Data NB K-Fold Cross Validation Dengan 10-Folds

Pada tabel 3.1 hasil akurasi yang diperoleh dengan penerapan metode *Naive Bayes* menggunakan teknik evaluasi *K-Fold Cross Validation* untuk klasifikasi dalam memprediksi keberhasilan *Immunotherapy Data Set* telah meningkat. Metode evaluasi *10-Fold Cross Validation* terbukti memberikan hasil akurasi lebih baik dengan nilai akurasi 81,11% dibandingkan metode evaluasi *5-Fold Cross Validation* dengan nilai akurasi 80,00%. Hasil yang diperoleh dari pengolahan kurva ROC seperti yang terlihat di Gambar 3.2 menghasilkan nilai AUC sebesar 0,63 yang termasuk kategori *good Classification*

b. Perbandingan Hasil Penelitian Dengan Penelitian Sebelumnya

Perbandingan hasil pada penelitian ini dengan penelitian-penelitian sebelumnya adalah untuk mengevaluasi hasil

keseluruhan pada penelitian ini. Penelitian ini adalah penelitian lanjutan dari penelitian-penelitian sebelumnya dengan objek data yang sama, yaitu *Immunotherapy Data Set*. Penelitian yang dilakukan oleh penelitian sebelumnya, dengan judul Analisis Algoritma Klasifikasi C.45 untuk Memprediksi Keberhasilan *Immunotherapy* Pada Penyakit Kutil dengan hasil akurasi sebesar 74,07%. dan penelitian selanjutnya dengan judul Penerapan Algoritma *Naive Bayes* Untuk Mengukur Nilai Akurasi Klasifikasi Diagnosis *Cryotherapy* Pada Penyakit Kutil dengan hasil akurasi adalah 91,67%.

Tabel 4. Perbandingan Hasil Penelitian

No	Judul Penelitian	Algoritma	Akurasi
1.	Analisis Algoritma Klasifikasi C.45 untuk Mempredik	Algoritma C.45	74,07%

	si Keberhasilan <i>Immunotherapy</i> pada penyakit kutil		
2.	Prediksi Keberhasilan <i>Immunotherapy</i> Pada penyakit kutil dengan Menggunakan Algoritma <i>Naive Bayes</i>	Algoritma <i>Naive Bayes</i>	81,11%

Pada Tabel 3.2 adalah perbandingan hasil penelitian pada objek *Immunotherapy Data Set* dari seluruh penelitian yang telah dilakukan. Dari tabel tersebut terbukti bahwa *Naive Bayes* dengan teknik *10-Fold Cross Validation* berhasil meningkatkan hasil akurasi pada klasifikasi deteksi *Penyakit Kutil* menjadi nilai akurasi terbaik dengan 81,11% dan ROC sebesar 0.63.

4. Kesimpulan

Pada penelitian ini dilakukan eksperimen terhadap algoritma *Naive Bayes* untuk mengklasifikasi *Immunotherapy Data Set*. Untuk mencari hasil optimal dari klasifikasi, dilakukan metode dalam validasi performa klasifikasi. metode *K-Fold Cross Validation*. Kesimpulan yang didapat dari hasil penelitian ini adalah:

- Telah diterapkan algoritma klasifikasi *Naive Bayes* pada *Immunotherapy Data Set* menggunakan metode *k-fold validasi* performa klasifikasi.
- Telah diketahui hasil akurasi dari algoritma *Naive Bayes* untuk klasifikasi *Immunotherapy Data Set* dengan hasil akurasi optimal sebesar 81,11% menggunakan *10-Fold Cross Validation*.

Saran-Saran

Penelitian ini hanya terbatas pada eksperimen algoritma *Naive Bayes* dan analisa terhadap hasil eksperimen tersebut,

dan tidak mengimplementasikan algoritma tersebut menjadi sebuah aplikasi.

- a. Pada penelitian berikutnya bisa dilakukan eksperimen menggunakan algoritma *Naive Bayes* menggunakan *dataset* lain untuk mengetahui performa *Naive Bayes Classifier* untuk *dataset* dengan jumlah data yang lebih beragam.
- b. Pada penelitian berikutnya bisa dilakukan eksperimen menggunakan *Immunotherapy Data Set* menggunakan metode algoritma *data mining* yang lain yang belum digunakan untuk *dataset* ini untuk mengetahui performa algoritma lainnya dalam mengklasifikasi penyakit kutil.
- c. Tingkat akurasi algoritma tersebut mencapai 81,11%. Dimana artinya pengujian data keberhasilan *Immunotherapy* dengan menggunakan algoritma *Naive bayes* 10 k-fold validation menghasilkan akurasi yang cukup baik dibandingkan penelitian sebelumnya yang dilakukan dengan data yang sama. Untuk penelitian berikutnya diperlukan penelitian lebih lanjut mengenai atribut yang mempengaruhi pengolahan data serta menggunakan lebih banyak data guna meningkatkan kualitas hasil penelitian.

Referensi

- Arifin, T. (2015). Metode Data Mining Untuk Klasifikasi Data Sel Nukleus Dan Sel Radang Berdasarkan Analisa Tekstur. *Informatika*, *II*(2), 425–433.
- Gorunescu, F. (2011). *Data mining: Concepts, models and techniques*. Retrieved from <https://doi.org/10.1007/978-3-642-19721-5>
- Han, J., Micheline, K., & Jian, P. (2012). *DATA MINING (Concept and Techniques)*. Retrieved from <https://doi.org/10.1017/CBO9781107415324.004>.
- Herliana, A., Arifin, T., Susanti, S., & Hikmah, A. B. (2018). Feature Selection of Diabetic Retinopathy Disease Using Particle Swarm Optimization and Neural Network. *2018 6th International Conference on Cyber and IT Service Management (CITSM)*, (Citsm), 1–4. <https://doi.org/10.1109/CITSM.2018.8674295>
- Hermawan, A., Sukma, A. R., & Halfis, R. (2019). *Analisis Algoritma Klasifikasi C*

4 . 5 Untuk Memprediksi Keberhasilan *Immunotherapy* Pada Penyakit Kutil. *V*(2), 155–160. <https://doi.org/10.31294/jtk.v4i2>

- Ramadhan, R., & Mahmudy, W. F. (2019). *Optimasi Bobot Awal Extreme Learning Machine menggunakan Algoritme Genetika untuk Klasifikasi Penanganan Human Papilloma Virus*. *3*(6), 5616–5623.
- Setiawati, D. (2014). *Human Papilloma Virus Dan Kanker Serviks*. 450–459.
- Supriyatna, A., & Mustika, W. P. (2018). *Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil*. (2), 152–161.