

Komparasi Algoritma *K-Nearest Neighbor* dan *Naive Bayes* pada Pengobatan Penyakit Kutil Menggunakan *Cryotherapy*

(Comparison of the *K-Nearest Neighbor* and *Naive Bayes* Algorithm in the Treatment of Warts Using *Cryotherapy*)

Herlambang Brawijaya¹, Samudi², Slamet Widodo³

^{1,3}*Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Kamal Raya no 18. Ringroad Barat. Cengkareng. Jakarta Barat*

²*Sistem Informasi, Sekolah Tinggi Manajemen Informatika Dan Komputer Nusa Mandiri
Jl. Damai No 8, Warung Jati Barat, Pasar Minggu*

¹herlambang.braw@gmail.com

²samudi.net@gmail.com

³slamet.smd@bsi.ac.id

Abstrak— Pengobatan penyakit kutil menggunakan *Cryotherapy* merupakan salah satu jenis pengobatan penyakit kutil yang direkomendasikan oleh beberapa pakar kesehatan. Metode yang digunakan dengan menggunakan nitrogen cair untuk pembekuan pada penyakit kutil. Dalam penelitian ini dilakukan komparasi pengujian model dengan menggunakan *K-Nearest Neighbor* dan *Naive Bayes* untuk prediksi pengobatan penyakit kutil. Dalam proses pengujiannya, peneliti menggunakan aplikasi *rapidminer* untuk mengolah data dan melakukan pengujian. Hasil pengujian yang telah dilakukan menunjukkan pengujian menggunakan model *K-Nearest Neighbor* (*K-NN*) didapat nilai akurasi terbaik adalah 90,00% dengan nilai AUC sebesar 0,500 sedangkan hasil pengujian menggunakan model *Naive Bayes* didapat nilai akurasi lebih kecil dibandingkan dengan model *K-NN* yaitu 86,67% dengan nilai AUC sebesar 0,932. Berdasarkan pengujian yang sudah dilakukan dapat disimpulkan bahwa model *K-Nearest Neighbor* memiliki tingkat akurasi lebih baik dibandingkan dengan model *Naive Bayes* dalam prediksi pengobatan penyakit kutil menggunakan *Cryotherapy*.

Kata-kata Kunci: Data Mining, *K-Nearest Neighbor*, *Naive Bayes*, Penyakit Kutil, *Cryotherapy*

Abstract— Treatment of warts using *Cryotherapy* is a type of wart disease treatment recommended by several health experts. The method used is using liquid nitrogen for freezing in wart diseases. In this study a model test was performed using *K-Nearest Neighbor* and *Naive Bayes* for research on the treatment of warts. In the testing process, researchers use the *rapidminer* application to process data and conduct testing. The results of tests that have been

carried out on testing using the *K-Nearest Neighbor* (*K-NN*) model get the best test value that is 90.00% with an AUC value of 0.500 while the test results using the *Naive Bayes* model get higher values with the *K-NN* model that is 86 , 67% with an AUC value of 0.932. Based on testing that has been done, it can be concluded that the *K-Nearest Neighbor* model has a better rating compared to the *Naive Bayes* model in predicting treatment of warts using *Cryotherapy*.

Keywords: Data Mining, *K-Nearest Neighbor*, *Naive Bayes*, wart disease, *Cryotherapy*

I. PENDAHULUAN

Penyakit kutil adalah masalah kesehatan kulit yang biasanya ditandai dengan munculnya benjolan kecil pada permukaan kulit, penyakit ini disebabkan oleh virus yaitu human papilloma virus (HPV). Penyebaran virus penyebab kutil bisa terjadi dengan mudah, contohnya hanya bersentuhan langsung dengan seseorang penderita kutil, tetapi tidak semua orang yang bersentuhan dengan virus hpv akan menimbulkan penyakit kutil. kekebalan tubuh (imunitas) masing-masing orang berbeda dan sangat berpengaruh dalam penularan penyakit ini, seseorang yang memiliki imunitas yang bermasalah akan rentan untuk terserang penyakit kutil [1].

Virus HPS masuk ke kulit melalui mikroabrasi dan menginfeksi sel basal. Siklus hidup HPV berkaitan dengan diferensi keratinosit baik pada fase produktif atau non produktif, fase non produktif meliputi pembentukan genome viral dalam jumlah sedikit sesuai

dengan tingkat pembelahan sel basal. Fase produktif mengikuti proses diferensiasi keratinosit dan virus mengalami replikasi dalam jumlah besar, mengepresikan late gen serta menghasilkan viral progeny. Terdapat berbagai cara untuk mengobati penyakit kutil (virus hpv) yaitu dengan melakukan vaksin, cryotherapy, dan imunoteraphy. Salah satu contoh untuk mengobati penyakit kutil dengan metode cryotherapy adalah teknik pengobatan terapi dengan berendam di dalam es atau air yang dingin selama kurang lebih 30 menit [2].

Dengan menggunakan es atau *Cold Bath* adalah beberapa cara yang dapat dipakai saat terapi, pada terapi ini menggunakan modalitas yang dapat menyerap jaringan suhu yang mengakibatkan penurunan penurunan suhu ringan yang melewati mekanisme konduksi. Efek yang terjadi pada pendinginan ini dapat terjadi tergantung jenis aplikasi terapi dingin yang dipakai, kalori yang diserap pada area local cedera sehingga terjadinya penurunan suhu adalah inti dari terapi dingin ini. Cryotherapy menyebabkan kerusakan jaringan melalui pembentukan Kristal es ekstra dan intraseluler, distrupsi membrane sel dan perubahan sirkulasi bagian kulit. Cryotherapy tidak menyebabkan keterlibatan sistemik, sehingga sesuai untuk wanita hamil. Terapi jenis ini sering tidak sesuai untuk lesi luas dan efek local yang sering terjadi adalah nekrosis, nyeri, pembentukan bula, edem dan hipopigmentasi. Pemakaian cryotherapy pada anak tidak dianjurkan. Cryotherapy tunggal mempunyai clearance rate sebesar 54-88% dengan tingkat rekuensi 21-40% [3].

Untuk menentukan klasifikasi pengobatan penyakit kutil dengan menggunakan cryotherapy terdapat banyak metode yang dapat digunakan seperti Naïve bayes, K-Nearest Neighbor, Decision Tree, K-Means, dan lain sebagainya, Namun untuk memilih metode yang paling cocok, dapat dilakukan komparasi antara beberapa metode. Dalam penelitian sebelumnya yang dilakukan oleh kelompok kami dengan menggunakan model algoritma Decision Tree menghasilkan akurasi sebesar 83,33%, metode K-NN sebesar 91,11%, dan metode Naïve Bayes sebesar 84,44%.

Pada pembahasan kali ini pada *machine learning* kami menggunakan dua metode yaitu Naïve Bayes dan K-Nearest Neighbour karena kedua metode ini memiliki akurasi yang tertinggi dibandingkan metode lain yang telah di uji, adapun ciri khas metode algoritma dalam proses klasifikasi ataupun prediksi, seperti penerapan pengaruh keberhasilan penyembuhan penyakit kutil dengan cryotherapy.

Machine learning dan algoritma data mining digunakan untuk menganalisa dataset dalam jumlah

yang banyak, menemukan dan mengekstrak pengetahuan dari dataset tersebut. Algoritma data mining dapat menganalisa kumpulan fakta dan informasi (data) untuk menentukan pola yang tidak diketahui dalam basis data (database) dalam jumlah banyak dari beberapa instansi seperti manufacturing, perbankan, asuransi, marketing, dan kesehatan. Umumnya algoritma data mining diterapkan untuk tujuan mengurangi biaya, meningkatkan kualitas penelitian, dan meningkatkan jumlah *income* bisnis [4].

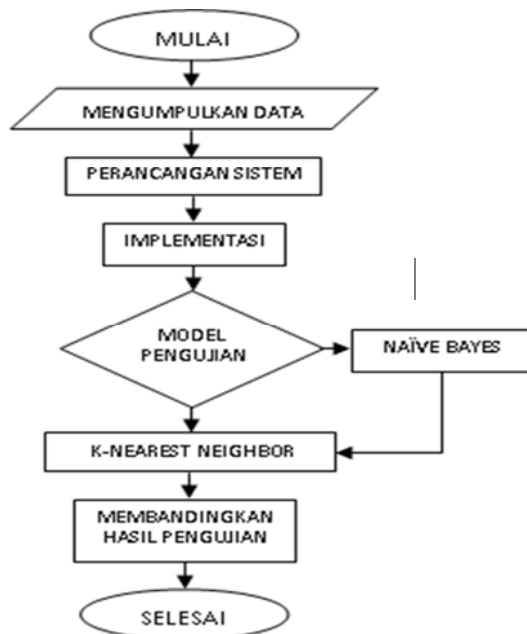
K-Nearest Neighbour dan Naïve bayes merupakan dua algoritma data mining yang digunakan untuk melakukan klasifikasi. Penelitian ini bertujuan untuk melakukan pengolahan dataset tentang penyembuhan penyakit kutil dengan metode cryotherapy dan menggunakan algoritma K-Nearest Neighbour dan Naïve Bayes dalam memprediksi keberhasilan pengobatan penyakit kutil dengan menggunakan metode cryoteraphy. Berdasarkan riset/penelitian yang dilakukan sebelumnya dinyatakan bahwa tingkat nilai akurasi yang dihasilkan algoritma metode K-Nearest Neighbour sangat tinggi dibandingkan dengan algoritma Naïve Bayes. Pada percobaan melakukan pengolahan data metode K-Nearest Neighbour rata-rata nilai akurasinya lebih unggul dibandingkan dengan menggunakan metode algoritma data mining yang lain.

Pada penelitian sebelumnya terdapat perbedaan tingkat akurasi beberapa metode untuk klasifikasi. Berdasarkan dari beberapa penelitian sebelumnya, metode K-Nearest Neighbor dan Naïve Bayes merupakan metode yang memiliki akurasi sangat tinggi. Untuk itu penelitian kami akan melakukan komparasi antara metode algoritma K-Nearest Neighbor dan Naïve Bayes untuk mengetahui metode mana yang lebih baik dalam membantu melakukan klasifikasi terhadap keberhasilan pengobatan penyakit kutil dengan menggunakan *cryotherapy*. Berdasarkan uraian deskripsi pendahuluan, maka judul yang digunakan untuk jurnal ini adalah "*Komparasi Algoritma K-Nearest Neighbor dan Naïve Bayes Untuk Memprediksi Keberhasilan Pengobatan Penyakit Kutil*".

II. METODE

Metode penelitian yang kami gunakan komparasi antara dua perbandingan model atau metode klasifikasi dengan menggunakan metode algoritma data mining yaitu metode K-Nearest Neighbor (K-NN) dan Naïve Bayes, dua jenis metode ini adalah bagian dari model atau metode supervised learning [5]. K-Nearest Neighbour (K-NN) adalah metode yang dikenal paling sederhana [5] sedangkan Naïve Bayes adalah sebuah

metode yang dapat menampilkan menggunakan label kelas terkait walaupun dengan data *training* yang sedikit [6]. Hasil dari setiap metode kemudian dicocokkan dengan *k-fold cross validation*. Gambar dari alur penelitian ini seperti pada Gambar 1.



Gambar 1. Metode penelitian

A. Algoritma K-Nearest Neighbor

Salah satu algoritma paling populer dalam *machine learning* adalah algoritma K-nearest neighbor (k-nn) dengan proses mudah dan sederhana [7]. Berdasarkan nilai dari variabel target yang terasosiasi dengan nilai variabel prediktor k-nn salah satu dari algoritma *supervised learning* dengan proses belajar. Dan dalam algoritma k-nn data harus memiliki label ketika data baru diberikan selanjutnya dibandingkan dengan data yang telah ada, diambil dari data yang mirip melihat dari data tersebut.

Berikut langkah-langkah algoritma k-NN:

- 1) Menentukan parameter K
- 2) Menghitung jarak diantara data uji dengan data latih, jika data tersebut berbentuk numerik harus menggunakan *Euclidean distance* pada persamaan 3.
- 3) Kemudian jarak diurutkan secara descending
- 4) Menentukan jarak terdekat pada parameter K
- 5) Jumlah kelas terbanyak di klasifikasikan

Algoritma K-Nearest Neighbor (KNN) adalah mencari range (jarak) yang paling dekat dengan k tetangga (neighbour) terdekat dalam data training dengan data yang akan diolah [8]. Teknik mengelompokkan

data baru dengan cara menghitung jarak data baru ke beberapa data/tetangga (neighbour) paling dekat. Algoritma K-Nearest Neighbour (KNN) merupakan *instead-based learning*. Dimana data training disimpan sehingga klasifikasi untuk tumpukan dataset baru (record) yang belum diklasifikasi dapat ditemukan dengan membandingkan kemiripan yang paling banyak dalam data training [9][10]. Permasalahan yang diketahui dimiliki dalam metode K-Nearest Neighbour (KNN) yaitu menemukan nilai terdekat K dari tetangga pada query dataset yang digunakan [11][12]. Untuk menghitung distance (jarak) dalam K-Nearest Neighbour (KNN) digunakan fungsi Eucliden Distance sebagaimana persamaan 1.

$$euc = \sqrt{\sum_i^n ((X_2)_i - (X_1)_i)^2} \quad \dots (1)$$

Keterangan :

- X_2 = Data latih
- X_1 = Data uji
- i = variable data
- n = dimensi data

B. Algoritma Naïve Bayes

Naïve bayes merupakan salah satu metode yang menghitung sekumpulan probabilitas dengan menjumlahkan frekuensi nilai atau data dari dataset yang diberikan [13]. *Naïve bayes* merupakan langkah-langkah dalam menyelesaikan masalah, klasifikasi yang sangat tepat, akurat, simple, dan efisiensi (proses penalaran dilakukan memanfaatkan input yang ada dengan cara yang relatif cepat). Langkah-langkah *naïve bayes* bertujuan untuk melakukan klasifikasi data terhadap kelas tertentu dan dapat menangani data baik yang bersifat diskrit maupun *continue* adalah kelebihan lain dari *naïve bayes* [14][15], seperti pada persamaan 2.

$$P(x_1, \dots, x_k | C) = P(x_1 | C) \times \dots \times P(x_k | C) \quad \dots (2)$$

Adalah nilai probabilitas yang diberikan pada persamaan 2. Ketika diberikan k atribut yang saling bebas adalah dalam proses mencari kelas terbaik.

Jika atribut ke-I bersifat diskrit atau kategori, maka $P(x_i | C)$ dapat diestimasi sebagai frekuensi relative sampel yang memiliki nilai x_i sebagai atribut ke-i dalam kelas C. maka $P(x_i | C)$ dapat dicari menggunakan *densitas gauss* (persamaan 3).

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \quad \dots (3)$$

Keterangan :

σ^2 = standar deviasi

μ = mean

Berikut contoh penerapan naïve bayes dengan menggunakan data training berjumlah 60 data training. Menghitung mean dan standar deviasi setiap fitur numeric adalah tahapan awal. Selanjutnya yaitu dengan menghitung probabilitas fitur kategori kemudian dengan menghitung probabilitas di setiap rekomendasi adalah tahapan selanjutnya.

C. Dataset

Pada Penelitian ini Dataset yang digunakan adalah dataset yang bersifat publik dan mengarah pada penelitian Khozeimeh [1] pengobatan kutil dengan metode cryotherapy, kemudian data-data di uji pada metode klasifikasi yang diterapkan aplikasi *rapidminer*. Sehingga dilanjutkan lagi dengan validasi data.

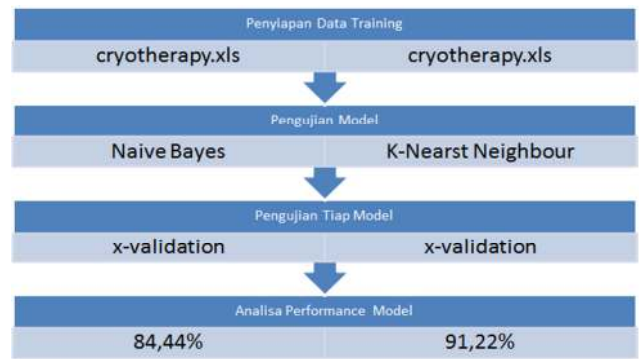
D. Akurasi

Pengujian akurasi merupakan suatu ukuran seberapa dekat (valid) hasil pengukuran terhadap angka sebenarnya. Akurasi dapat diperoleh dari persentase kebenaran, yaitu perbandingan antara jumlah data benar dengan keseluruhan dataset seperti pada persamaan 4.

$$akurasi = \frac{\text{total data benar}}{\text{total data}} \times 100\% \quad \dots(4)$$

III. HASIL DAN PEMBAHASAN

Pada penelitian pengobatan kutil dengan metode *cryotherapy* dibagi menjadi dua model pengujian adalah pengujian dengan model k-nearest neighbor (k-nn) dan model naïve bayes, kemudian hasil di klasifikasi dari masing masing data yang di ambil dan dilakukan dengan proses validasi data. Kemudian hitung standar deviasi dari tiap-tiap metode. sistem ini menggunakan *machine learning Rapid Miner*. Gambar 2 adalah kerangka kerja implementasi sistem menggunakan *Rapid Miner*.



Gambar 2. Kerangka kerja implementasi sistem menggunakan *rapidminer*

A. Pengujian nilai K dengan K-Nearest Neighbor

Pengujian dengan K-Nearest Neighbor (K-NN) dilakukan dengan cara menginisialisasi nilai K, pada pengujian ini dilakukan sebanyak 15 kali mengubah isi nilai K dalam setiap skenario pengujian K dilakukan sejumlah 8 record pengujian data training. Hasil dari setiap pengujian dengan nilai k terbaik akan digunakan dalam pengujian gabungan (komparasi) metode *Naïve* dan bayes K-Nearest Neighbor (K-NN). Nilai K yang diimplementasikan pada pengujian komparasi nanti hanya menggunakan angka terkecil dengan akurasi tertinggi dikarenakan untuk mempermudah perhitungan komparasi selanjutnya. Data training yang akan diolah pada pengujian sebanyak 60 data dan data testing sebanyak 30 data. Hasil pengujian berdasarkan nilai atribut K adalah sebagai berikut:

K1	: 90,00%
K3	: 68,33%
K5	: 70,00%
K7	: 65,00%
K9	: 63,00%
K11	: 65,00%
K13	: 51,67%
K15	: 63,33%

Dikarenakan K1 telah mendapatkan tingkat akurasi yang sangat tinggi dengan nilai K lain yaitu 90,00%, maka peneliti mengambil K1 sebagai bahan penelitian kami dengan 60 data *training* dan 30 data *testing* dari 90 jumlah *record dataset*.

Pada pengujian berdasarkan nilai atribut K data yang digunakan sebanyak 90 record dataset yang terbagi menjadi 60 data training dan 30 data testing. Hasil pengujian terendah terjadi ketika nilai K yang bernilai 9 yaitu 63,00%. Hasil pengujian yang tertinggi dan terbaik pada nilai atribut K yang bernilai 1 yaitu 90,00%. Dapat ditarik kesimpulan bahwa akurasi K-Nearest Neighbor akan dipengaruhi terhadap jumlah

nilai atribut K. Semakin banyak nilai atribut K, maka semakin rendah tingkat akurasi, hal ini dikarenakan oleh atribut yang digunakan memiliki kesamaan dalam jumlah banyak sehingga semakin banyak tetangga atau nilai K yang diambil semakin banyak data dari kelas lain ikut dijadikan pertimbangan keputusan. Pada pengujian yang dilakukan oleh metode K-Nearest Neighbor (K-NN) akurasi yang tertinggi diperoleh pada saat nilai atribut K bernilai 1 yang nantinya akan digunakan dalam pengujian komparasi metode K-Nearest Neighbor dan Naïve Bayes.

B. Pengujian performance tiap model

Di tahap ini peneliti mulai membandingkan tingkat akurasi dari tiap model yaitu K-Nearest Neighbor dan

Naïve Bayes dengan dataset yang sama yaitu 60 *training* set dan 30 *testing* set, disini kami memperoleh untuk model Naïve Bayes dengan nilai performance 86,67% lebih rendah bila dibandingkan menggunakan K-NN yang mendapatkan hasil *performance* 90,00% dengan jumlah dataset yang sama. Dua model klasifikasi yang memiliki ciri khas yang berbeda masing-masing model sehingga dalam proses klasifikasi memiliki langkah-langkah yang berbeda pula atau sering disebut dengan algoritma yang akan diterapkan dalam melakukan klasifikasi data pengobatan penyakit kutil dengan metode cryotherapy. Berikut hasil dari penerapan dengan menggunakan model Naïve Bayes (Gambar 3).

Table View Plot View

accuracy: 86.67% +/- 10.00% (mikro: 86.67%)

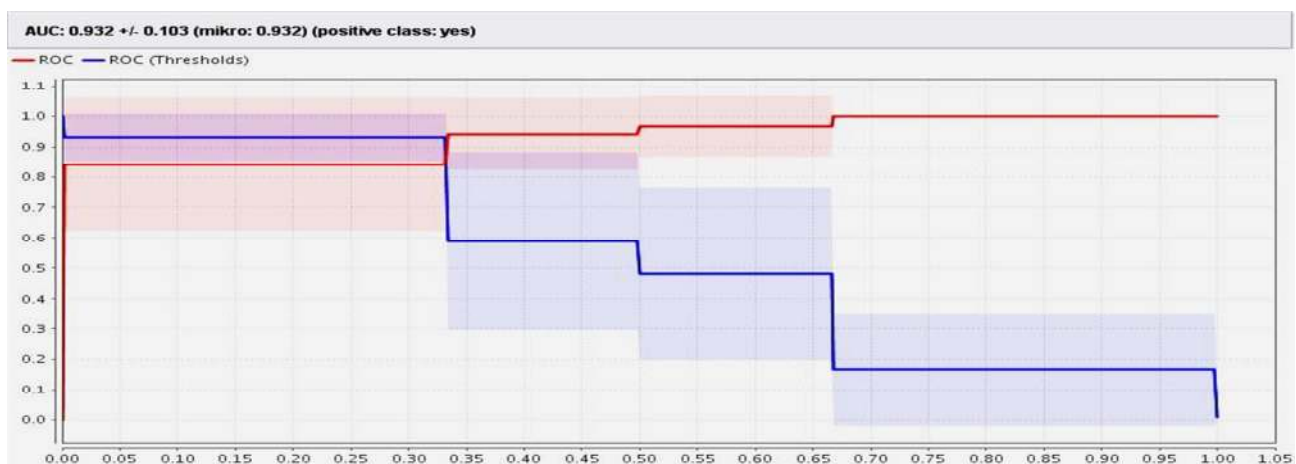
	true No	true yes	class precision
pred. No	22	2	91.67%
pred. yes	6	30	83.33%
class recall	78.57%	93.75%	

Gambar 3. Performance Naïve Bayes

Dengan hasil yang ditunjukkan pada Gambar 3 dapat dilihat bahwa model Naïve Bayes dengan *training set* 66,67% dari jumlah data 90 data dan 30 yang dinyatakan (*yes*) diprediksi dengan benar dan sebanyak 26 dinyatakan (*no*) diprediksi dengan benar pula, sedangkan 2 data yang menyatakan (*yes*) dengan prediksi yang salah, dan sebanyak 6 data yang (*no*) juga diprediksikan salah. Maka didapatkan *Accuracy* yang

diterapkan *Naïve Bayes* adalah 86,67%, Sensitivity 91,67%, Specificity 83,33%, PPV 78,57%, NPV 93,75%.

Berikut hasil AUC dari model Naïve Bayes (Gambar 4).



Gambar 4. AUC Naïve Bayes

Terlihat pada Gambar 4 adalah AUC (*Area Under Curve*) untuk data *training* yaitu sebesar 0,932 dengan nilai akurasi *Excelent Accuracy*.

Berikut hasil pengujian yang dilakukan menggunakan model *K-Nearest Neighbor* (Gambar 5).

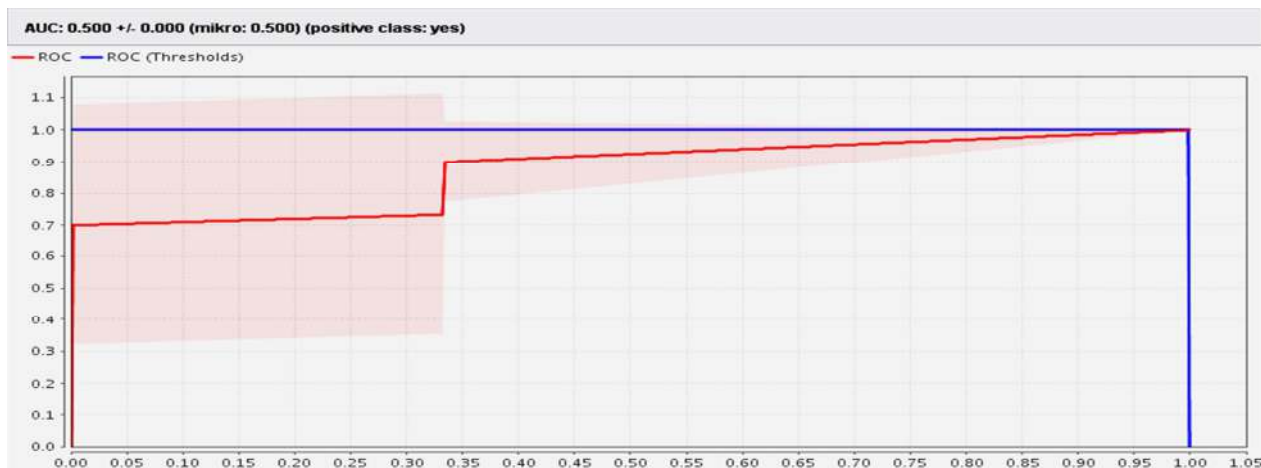
accuracy: 90.00% +/- 11.06% (mikro: 90.00%)			
	true No	true yes	class precision
pred. No	26	4	86.67%
pred. yes	2	28	93.33%
class recall	92.86%	87.50%	

Gambar 5. Performance K-NN

Dengan hasil yang ditunjukkan pada Gambar 5 dapat dilihat bahwa *accuracy* yang diterapkan K-Nearest Neighbor sebesar 90,00%, Sensitivity 86,67%, Specificity 93,33%, PPV 92,85%, NPV 87,50%.

Berikut adalah grafik ROC terlihat AUC (*Area Under Cover*) yang dihasilkan oleh model K-NN sebesar 0,500 dengan nilai akurasi yang terbilang medium atau rata-rata (Gambar 6). Hal tersebut dipengaruhi oleh jika semakin banyaknya data *training*

yang digunakan maka semakin baik dan semakin lengkap juga model klasifikasinya yang dibentuk berdasarkan fakta dari data *training* yang diolah. Sehingga ketika saat melakukan prediksi pada klasifikasi data *testing* atau data yang baru, maka akurasi yang didapatkan akan semakin baik atau semakin tinggi.



Gambar 6. AUC K-NN

IV. PENUTUP

Berdasarkan pengujian dari dua metode machine learning yang telah diterapkan pada penelitian kami, memperoleh rata-rata akurasi system dengan penyembuhan metode *crtiotherapy* ketika menggunakan k-nearest neighbor (k-nn) sebesar 90,00%. Sedangkan ketika menggunakan metode *naïve bayes* dihasilkan rata-rata akurasi sebesar 86,67%, namun jika diperhatikan ROC nya AUC K-NN lebih tinggi jika dibandingkan dengan AUC *Naïve Bayes*, 0,500 : 0,932 dengan demikian setelah diterapkan

menggunakan model k-nearest neighbor (k-nn) nilai akurasi dalam melakukan klasifikasi lebih baik dibandingkan dengan *naïve bayes*. Selain itu, jarak akurasi setiap eksperimen dengan rata-rata akurasi lebih dekat saat menggunakan k-nearest neighbor (k-nn) dibandingkan dengan *naïve bayes* hal ini sesuai dengan nilai standar deviasi yang dihasilkan dari masing-masing metode.

DAFTAR PUSTAKA

- [1] F. Khozeimeh *et al.*, “Intralesional immunotherapy compared to cryotherapy in the treatment of warts,” *Int. J. Dermatol.*, vol. 56, no. 4, pp. 474–478, 2017.
- [2] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layegh, and S. Nahavandi, “An expert system for selecting wart treatment method,” *Comput. Biol. Med.*, vol. 81, pp. 167–175, 2017.
- [3] H. Amalia and E. Evicienna, “Komparasi Metode Data Mining Untuk Penentuan Proses Persalinan Ibu Melahirkan,” *J. Sist. Inf.*, vol. 13, no. 2, p. 103, 2017.
- [4] N. Saputra, T. B. Adji, and A. E. Permanasari, “Analisis Sentimen Data Presiden Jokowi dengan Preprocessing Normalisasi dan Stemming menggunakan Metode Naive Bayes dan SVM,” *J. Din. Inform.*, vol. 5, no. 1, pp. 1–12, 2015.
- [5] P. Piro, R. Nock, F. Nielsen, and M. Barlaud, “Leveraging k-NN for generic classification boosting,” *Neurocomputing*, vol. 80, pp. 3–9, 2012.
- [6] D. Kurnianingtyas, B. A. Rahardian, D. P. Mahardika, A. K. A., and D. A. K., “Sistem Pendukung Keputusan Diagnosis Penyakit Sapi Potong Menggunakan K-Nearest Neighbour (K-NN),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 2, p. 122, 2017.
- [7] P. Harrington, *Machine Learning in Action*, vol. 37, no. 3, 2012.
- [8] M. M. Jain and P. V. Richariya, “An Improved Techniques Based on Naive Bayesian for Attack Detection,” *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 1, pp. 324–331, 2012.
- [9] Yeni Kustiyahningsih and N. Syafa’ah, “Sistem Pendukung Keputusan Untuk Menentukan Jurusan Pada Siswa Sma Menggunakan Metode Knn Dan Smart,” *JSII*, vol. 1, no. 1, pp. 19–28, 2014.
- [10] F. Gorunescu, “Data mining: Concepts, models and techniques,” *Intell. Syst. Ref. Libr.*, vol. 12, 2011.
- [11] Y. C. Liaw, M. L. Leou, and C. M. Wu, “Fast exact k nearest neighbors search using an orthogonal search tree,” *Pattern Recognit.*, vol. 43, no. 6, pp. 2351–2358, 2010.
- [12] Y. C. Liaw, C. M. Wu, and M. L. Leou, “Fast k-nearest neighbors search using modified principal axis search tree,” *Digit. Signal Process. A Rev. J.*, vol. 20, no. 5, pp. 1494–1501, 2010.
- [13] Bustami, “Penerapan Algoritma Naive Bayes untuk Mengklasifikasi Data Nasabah,” *TECHSI J. Penelit. Tek. Inform.*, vol. 4, pp. 127–146, 2010.
- [14] N. D. Prayoga, N. Hidayat, and R. K. Dewi, “Sistem Diagnosis Penyakit Hati Menggunakan Metode Naïve Bayes,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 8, pp. 2666–2671, 2018.
- [15] J. Wu and Z. Cai, “Attribute weighting via differential evolution algorithm for attribute Weighted Naive Bayes (WNB),” *J. Comput. Inf. Syst.*, vol. 7, no. 5, pp. 1672–1679, 2011.

