

Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Algoritma *Naïve Bayes* Dan *Iterative Dichotomizer 3 (ID3)*

Putri Pratiwi¹, Dima Dwifa², Anita Desiani^{*3}, Ali Amran⁴, Bambang Suprihatin⁵

Jurusan Matematika Universitas Sriwijaya, Indralaya
Jl. Raya Palembang - Prabumulih No.Km. 32, Indralaya Indah, Kec. Indralaya, Kabupaten Ogan Ilir,
Sumatera Selatan 30862

¹putri0421pr@gmail11.com

²dimadwifa96@gmail.com

^{*3}anita_desiani@unsri.ac.id

⁴ali_amran@mipa.unsri.ac.id

⁵bambang@unsri.ac.id

Intisari — Kanker paru-paru merupakan penyakit yang disebabkan oleh keganasan tumor dari bronkus. Pada tahun 2020 terdapat kasus kematian mencapai 30.843 jiwa. Agar menghambat angka kematian akibat kanker paru-paru, diperlukan alat untuk deteksi dini akibat kanker paru-paru. Pada penelitian ini menggunakan dataset kanker paru-paru yang memiliki 309 data. Teknik uji yang digunakan pada dataset kanker paru-paru ini adalah *percentage split* dan *k-fold cross validation* dengan algoritma naive bayes dan *iterative dichotomizer 3 (ID3)*. Parameter penilaian yang digunakan untuk menentukan algoritma terbaik adalah akurasi, presisi dan *recall*. Hasil penelitian yang menggunakan teknik uji *percentage split*, diperoleh nilai akurasi, presisi dan *recall* tertinggi pada algoritma naive bayes yaitu akurasi sebesar 87%, presisi sebesar 91% dan *recall* sebesar 94% untuk kelas YES (*positif lung cancer*). Penelitian yang dilakukan dengan teknik uji menggunakan *k-fold cross validation* memberikan nilai terbaik pada algoritma ID3 dengan nilai akurasi 92%, presisi sebesar 94% dan *recall* sebesar 97% untuk kelas YES (*positif lung cancer*). Dengan demikian, penelitian yang dilakukan dengan metode *k-fold cross validation* memiliki nilai yang lebih tinggi dibandingkan nilai menggunakan teknik uji *percentage split*. Hal tersebut memberikan kesimpulan bahwa pada penelitian ini, diperoleh algoritma terbaik yaitu ID3 dengan teknik uji *k-fold cross validation* pada dataset kanker paru-paru.

Kata Kunci — Data mining, Kanker paru-paru, Naive bayes, *Iterative dichotomizer 3 (ID3)*, *percentage split* dan *k-fold cross validation*

Abstract — Lung cancer is a disease caused by malignant tumors of the bronchi. In 2020 there were 30,843 deaths. In order to inhibit the death rate from lung cancer, tools are needed for early detection due to lung cancer. This study used a lung cancer dataset that had 309 data. The test techniques used in this lung cancer dataset are percentage split and k-fold cross validation with naïve bayes and iterative dichotomizer 3 (ID3) algorithms. The scoring parameters used to determine the best algorithm are accuracy, precision and recall. The results of the study using the percentage split test technique, obtained the highest accuracy, precision and recall values on the naïve bayes algorithm, namely accuracy of 87%, precision of 91% and recall of 94% for the YES (positive lung cancer) class. Research conducted using k-fold cross validation test technique gave the best value on the ID3 algorithm with an accuracy value of 92%, precision of 94% and recall of 97% for the YES (positive lung cancer) class. Thus, research conducted with the k-fold cross validation method has a higher value than the value using the percentage split test technique. This concludes that in this study, the best algorithm was obtained, namely ID3 with the k-fold cross validation test technique on lung cancer datasets.

Keywords — Data mining, Lung cancer, Naive bayes, *Iterative dichotomizer 3 (ID3)*, *percentage split* and *k-fold cross validation*.

I. PENDAHULUAN

Kanker paru-paru adalah kondisi paru-paru yang melemah akibat adanya pertumbuhan sel yang tidak terkendali [1]. Penyakit kanker paru-paru berasal dari keganasan tumor yang asalnya dari bronkus. Pada tahun 2020, angka

penyakit kanker paru-paru mencapai 34.783 kasus dan mengakibatkan 30.843 kematian [2].

Gejala umum yang dialami seorang penderita kanker paru-paru yaitu nyeri pada bagian dada ketika sedang batuk dan bernafas [3]. Terkadang penderita kanker paru-paru menganggap gejala tersebut merupakan gejala pada penyakit batuk biasa dan tidak perlu

dilakukan pemeriksaan lebih lanjut [4]. Penyebab lainnya adalah petugas layanan kesehatan yang tidak melakukan pemeriksaan ulang kepada pasien [5]. Oleh karena itu, untuk mengurangi resiko dari penyakit kanker paru-paru perlu dilakukan pemeriksaan sejak dini. Salah satu cara untuk mengidentifikasi sejak dini adalah dengan menggunakan beberapa algoritma yang ada pada *machine learning* [6]. Algoritma yang sering digunakan yaitu algoritma naïve bayes dan *iterative dichotomizer 3* (ID3).

Algoritma naïve bayes merupakan suatu algoritma dari pengembangan teorema bayes [7][8]. Kelebihan naïve bayes ialah tidak tergantung pada distribusi data dikarenakan menganggap semua data seragam [9]. Beberapa penelitian sebelumnya yang telah melakukan klasifikasi dengan algoritma Naive bayes dalam bidang kesehatan adalah klasifikasi kanker payudara (H. Oktavianto dan R. P. Handri, 2020 [10]) yang memberikan akurasi sebesar 96.9%, klasifikasi penyakit diabetes mellitus (A. Ridwan, 2020 [11]) dengan hasil akurasi 90.20%, dan klasifikasi penyakit stroke (A. Byna dan M. Basit, 2020 [12]) dengan hasil akurasi 97.6%. Kekurangan algoritma Naive bayes adalah cenderung lebih baik dalam menangani atribut dengan tipe data kategorik.

Algoritma lain yang dapat digunakan untuk klasifikasi dataset kanker paru-paru adalah *iterative dichotomizer 3* (ID3). Algoritma ID3 membangun suatu keputusan melalui pohon dan percabangan dengan menggunakan perhitungan nilai *entropy* dan *informasi gain* [13]. Kelebihan dari algoritma ID3 adalah dapat menangani atribut dengan tipe data kategorik maupun numerik dengan baik, daerah pengambilan keputusannya dapat diubah dari kompleks dan sangat *general* menjadi simple dan spesifik [14]. Beberapa penelitian sebelumnya yang menggunakan algoritma ID3 adalah klasifikasi kesehatan kehamilan (H. Hikmatulloh *et.al*, 2019 [15]) yang memberikan akurasi sebesar 80.33%, klasifikasi kesuburan pria (K. Ma'mur, 2019 [16]) dengan hasil akurasi 92%, dan klasifikasi penyakit diabetes mellitus tipe 2 (A. E. Tyasti *et.al*, 2015 [17]) dengan hasil akurasi 78.5%.

Salah satu dataset kanker paru-paru yang sering digunakan ialah dataset kanker paru-paru dari situs kaggle <https://www.kaggle.com/datasets/mysarahmadbhat/lung-cancer>. Pada dataset kanker paru-paru jumlah data per label tidak seimbang, dimana data yang mengalami kanker paru-paru sebanyak 270 data dan yang tidak mengalami kanker paru-paru sebanyak 39 data. Data yang tidak seimbang akan mempengaruhi kinerja klasifikasi karena kurangnya ketersediaan informasi mengenai data yang jumlahnya minoritas [18]. Dalam melakukan klasifikasi terdapat dua teknik uji yang sering digunakan yaitu *percentage split* dan *k-fold cross validation* [19]. *Percentage split* memiliki kelebihan yaitu mudah digunakan dan tidak memerlukan waktu *training* yang lama [20]. Kekurangan penggunaan *percentage split* ialah pada data yang tidak seimbang ketika dilakukan pemotongan data terdapat data yang tidak ikut terpotong dan mengakibatkan informasi pembelajaran *machine learning* berkurang. Salah satu cara mengatasi kekurangan *percentage split* adalah dengan menggunakan teknik uji *k-fold cross validation* [21].

K-fold cross validation adalah teknik *training* data yang membagi data secara berulang [22]. Misalnya data dibagi menjadi 10, maka data akan mengalami pergantian sebagai data *training* maupun data *testing*. Kelebihan dari *k-fold cross validation* adalah dapat bekerja dengan rentang yang luas [23]. Hal ini dikarenakan setiap data akan mengalami proses *training* dan *testing* secara bergantian. Sehingga untuk data yang tidak seimbang informasi masih tetap dapat diperbaiki.

Pada penelitian ini, akan melakukan klasifikasi terhadap penyakit kanker paru-paru dengan menerapkan algoritma naïve Bayes dan ID3. Dalam mengetahui bagaimana kinerja dari kedua algoritma ini akan digunakan dua teknik uji yaitu *percentage split* dan *k-fold cross validation*. Penelitian dilakukan untuk mengetahui mana penerapan teknik uji yang paling baik pada algoritma naïve bayes dan ID3 akan diukur hasil uji berdasarkan *recall*, presisi, dan akurasi pada setiap teknik uji.

II. METODOLOGI PENELITIAN

A. Deskripsi Data

Dataset yang digunakan pada penelitian ini bersumber pada situs kaggle

<https://www.kaggle.com/datasets/mysarahma/dbhat/lung-cancer>. Dataset terdiri dari 309 data dengan 16 atribut. Atribut-atribut yang digunakan pada dataset kanker paru-paru ditunjukkan pada Tabel 1.

Tabel 1. Atribut, Tipe Data dan Keterangan Atribut

Nama Atribut	Tipe Data	Keterangan
<i>Gender</i> (Jenis Kelamin)	Nominal	F: <i>Female</i> , M: <i>Male</i>
<i>Age</i> (Umur)	Numerik	Usia 47-87 tahun
<i>Smoking</i> (Merokok)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Yellow Fingers</i> (Jari Kuning)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Anxiety</i> (Kecemasan)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Peer Pressure</i> (Tekanan Sosial)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Chronic Disease</i> (Penyakit Kronis)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Fatigue</i> (Sakit Kepala)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Allergy</i> (Alergi)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Wheezing</i> (Mengi)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Alcohol Consuming</i> (Mengonsumsi Alkohol)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Coughing</i> (Batuk)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Shortness of Breath</i> (Sesak Nafas)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Chest Pain</i> (Nyeri dada)	Nominal	1: <i>NO</i> , 2: <i>YES</i>
<i>Lung Cancer</i> (kanker paru-paru)	Nominal	<i>YES</i> , <i>NO</i>

Berdasarkan tabel 1, dataset kanker paru-paru terdiri 15 atribut gejala dan satu atribut merupakan atribut kelas. Dataset kanker paru-paru memiliki dua kelas yaitu *YES* dan *NO*.

B. Praproses Data

Pada dataset kanker paru-paru terdapat data yang tidak seimbang yaitu 270 data *YES* (*positif lung cancer*) dan 39 data *NO* (*negatif lung cancer*) yaitu atribut *smoking*, *yellow fingers*, *anxiety*, *peer pressure*, *chronic disease*, *fatigue*, *allergy*, *wheezing*, *alcohol consuming*, *coughing*, *shortness of breath*, dan *chest pain*. Pada data yang tidak seimbang perlu dilakukan metode normalisasi. Metode yang digunakan untuk permasalahan nilai yang terlampaui jauh adalah *min-max normalization*. Rumus umum *min max normalization* dapat ditunjukkan pada Persamaan (1).

$$\text{normalized}(x) = \frac{\text{minrange}(x) + (x - \text{minvalue})(\text{maxrange} - \text{minrange})}{\text{maxvalue} - \text{minvalue}} \quad (1)$$

Pengujian dan pelatihan data dilakukan dengan menggunakan metode *percentage split* dan metode *k-fold cross validation*. Pada metode *percentage split* data dibagi menjadi data *training* 80% dan data *testing* 20%. Kemudian, pada metode *k-fold cross validation* nilai *k* yang digunakan sebesar 10. Data terbagi menjadi 9 kelompok data latih dan 1 kelompok data uji sehingga menjadi 10 kelompok data.

C. Pengolahan Data

1) Implementasi Algoritma Naïve Bayes

Pada pengklasifikasian dengan algoritma Naïve Bayes dilakukan beberapa langkah sebagai berikut [24]:

- Membaca data *training*
Langkah awal untuk menentukan data yang akan dianalisis yaitu dengan membaca data *training*. Pada penelitian ini, terdapat 8 data *training*/sampel.
- Menghitung peluang setiap kelas.
Tahap awal dari perhitungan algoritma naïve bayes ialah menghitung nilai peluang setiap kelas. Kelas yang digunakan yaitu “*YES*” (*positif lung*

cancer) dan “NO” (negatif lung cancer). Perhitungannya dilakukan dengan Persamaan 2 berikut:

$$P(kelas) = \frac{n(kelas)}{n(S)} \quad (2)$$

dimana $P(kelas)$ ialah peluang setiap kelas, dan $n(kelas)$ adalah banyaknya data (kelas) dari total data *training* serta $n(S)$ ialah banyaknya seluruh data.

- Menghitung peluang setiap atribut. Perhitungannya dilakukan dengan membandingkan atribut dari data *testing* dengan atribut dari data *training*.

$$P(X|kelas) = \frac{P(X \cap kelas)}{P(kelas)} \quad (3)$$

dengan $P(X|kelas)$ ialah peluang setiap atribut dan $P(X \cap kelas)$ adalah peluang atribut dengan kelas pada data *training* serta $P(kelas)$ ialah peluang kelas.

- Menghitung peluang akhir setiap kelas. Perhitungan peluang akhir setiap kelas menggunakan data *training* yang nilainya sudah diubah sesuai setiap atribut. Kemudian setiap atribut dikalikan dengan nilai peluang kelas. Jika sudah diperoleh kedua hasil dari setiap kelas, maka bandingkan kedua nilai. Lalu, yang bernilai paling tinggi, itulah hasilnya. Begitu sebaliknya.

2) Implementasi Iterative Dechotomizer 3 (ID3)

Tahapan dalam perhitungan algoritma *Iterative Dechotomizer* (ID3) yaitu sebagai berikut [15]:

- Menghitung nilai *entropy*. Nilai *entropy* adalah ukuran dari keseluruhan energi yang tidak dapat diubah menjadi usaha. Rumus untuk mendapatkan nilai *entropy* terdapat pada Persamaan 4.

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2 p_i \quad (4)$$

dimana, S adalah himpunan dataset, n merupakan jumlah partisi S , dan p_i

merupakan probabilitas yang didapat dari jumlah kelas dibagi total kasus.

3) Menghitung nilai Information Gain

Information gain digunakan untuk melakukan ranking nilai dari suatu data dengan penilaian data yang paling kuat adalah suatu kumpulan data. Rumus untuk mendapatkan *information gain* dapat dilihat pada Persamaan 5.

$$GAIN(S, A) = Entropy(S) - \sum_{V \in values(A)} \frac{S_V}{S} Entropy(S_V) \quad (5)$$

dimana, A merupakan atribut yang digunakan dan ada pada dataset, V adalah nilai mungkin untuk atribut A , $values(A)$ merupakan himpunan nilai mungkin untuk atribut A dan $Entropy(S_V)$ merupakan *Entropy* untuk sampel-sampel yang mempunyai nilai v .

- Menentukan *root node* (node akar). *Root* diperoleh dengan menentukan nilai informasi gain tertinggi dari atribut yang digunakan. Atribut yang telah dipilih sebagai akar tidak dihitung kembali *entropy* dan informasi gainnya.
- Membuat node cabang. Node cabang didapat dari perhitungan *entropy* dan informasi gain.

D. Analisis Hasil

Pada proses ini, *confusion matrix* digunakan untuk menampilkan hasil evaluasi berbentuk tabel matriks. *Class* pertama dianggap positif dan *class* kedua dianggap negatif. Evaluasi dengan *confusion matrix* menghasilkan nilai akurasi, presisi, *recall* dan *F1-Score* [25]. *Confusion matrix* adalah tabel yang menyatakan klasifikasi total data uji yang diprediksi benar dan salah [26]. Bentuk umum *confusion matrix* untuk klasifikasi biner ditunjukkan pada Tabel 2 berikut [27].

Tabel 2. Tabel *Confusion Matrix*

Class		Prediction	
		Positive	Negative
Actual	Positive	TP (True Positive)	FN (False Negative)
	Negative	FP (False Positive)	TN (True Negative)

dimana, TP (*True Positive*) ialah total *instance* positif diklasifikasikan positif, FN (*False Negative*) adalah total *instance* positif diklasifikasikan negatif, FP (*False Positive*) ialah total *instance* negatif diklasifikasikan positif, serta TN (*True Negative*) adalah total *instance* negatif diklasifikasikan positif.

Berdasarkan *confusion matrix* dapat diukur tingkat evaluasinya misalkan nilai akurasi, presisi, dan *recall* [28]. Akurasi ialah nilai untuk memprediksi benar atau salah berdasarkan kumpulan dari data. Presisi adalah informasi yang didapat oleh pengguna berupa tingkat ketepatan jawaban yang diberikan oleh sistem. *Recall* merupakan tingkat keberhasilan sistem dalam

mendapatkan kembali informasi. Berdasarkan [29][30][26] rumus untuk menghitung performa *confusion matrix* dapat dilihat pada Persamaan (6), Persamaan(7), dan Persamaan (8).

$$Akurasi = \frac{TP+TN}{TP+TN+FN+FP} \quad (6)$$

$$Presisi = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

Tabel 3. *Confussion Matrix* Pada Algoritma Naive Bayes

Kelas		Aktual teknik uji <i>percentage split</i>		Aktual teknik uji <i>k-fold cross validation</i>	
		YES (<i>positif lung cancer</i>)	NO (<i>Negatif Lung Cancer</i>)	YES (<i>positif lung cancer</i>)	NO (<i>Negatif Lung Cancer</i>)
Prediksi	YES (<i>positif lung cancer</i>)	51	3	257	13
	NO (<i>Negatif Lung Cancer</i>)	5	3	17	22

III. HASIL DAN PEMBAHASAN

A. Algoritma Naive Bayes

Teknik uji data yang digunakan adalah *percentage split* dan *k-fold cross validation*. Pada *percentage split* dibagi menjadi data training sebesar 80% dan data testing sebesar 20%. Metode *k-fold cross validation* data dibagi menjadi 10. *Confussion matrix* hasil penerapan naïve bayes dengan teknik *percentage split* dan *k-fold cross validation* dapat dilihat pada Tabel 3. Dari Tabel 3, dapat dilihat bahwa uji coba dengan algoritma naïve bayes memberikan hasil 51 kasus YES (*positif lung cancer*) diprediksi YES (*positif lung cancer*), 3 kasus YES (*positif lung cancer*) diprediksi NO (*negatif lung cancer*), 5 kasus NO (*negatif lung cancer*) diprediksi YES (*positif lung cancer*), dan 3 kasus NO (*negatif lung cancer*) diprediksi NO (*negatif lung cancer*). Selain itu, dari Tabel 3 dapat dihitung nilai akurasi diperoleh dari total data yang berhasil diprediksi dengan tepat sebesar 87%. Hasil pengukuran *precision* untuk kelas YES (*positif lung cancer*) sebesar 91% dan untuk kelas NO (*negatif lung cancer*) sebesar 50%.

Hasil pengukuran *recall* untuk kelas YES (*positif lung cancer*) yaitu 94% dan untuk kelas NO (*negatif lung cancer*) yaitu 38%. Teknik uji *k-fold cross validation* dengan algoritma Naive bayes memberikan hasil 257 kasus YES (*positif lung cancer*) diprediksi YES (*positif lung cancer*), 13 kasus YES (*positif lung cancer*) diprediksi NO (*negatif lung cancer*), 17 kasus NO (*negatif lung cancer*) diprediksi YES (*positif lung cancer*), dan 22 kasus NO (*negatif lung cancer*) diprediksi NO (*negatif lung cancer*). Selain itu, dari Tabel 5 dapat dihitung nilai akurasi berdasarkan data yang berhasil diprediksi dengan tepat sebesar 90%. Hasil pengukuran *precision* untuk kelas YES (*positif lung cancer*) adalah sebesar 94% dan untuk kelas NO (*negatif lung cancer*) adalah sebesar 63%. Hasil pengukuran *recall* untuk kelas YES (*positif lung cancer*) yaitu 95% dan untuk kelas NO (*negatif lung cancer*) yaitu 62%. Secara lebih lengkap, hasil presisi, *recall* dan akurasi dengan teknik *percentage split* dan *k-fold cross validation* dapat dilihat pada Tabel 4.

Tabel 4. Hasil Permodelan Algoritma Naive Bayes

Target Class	Percentage Split			K-fold cross validation		
	Presisi (%)	Recall (%)	Akurasi (%)	Presisi (%)	Recall (%)	Akurasi (%)
YES (<i>positif lung cancer</i>)	91	94	87	94	95	90
NO (<i>Negatif Lung Cancer</i>)	50	38		63	56	

Dari Tabel 4, diketahui bahwa algoritma naive bayes memiliki nilai *recall* dan akurasi lebih baik dengan menggunakan teknik uji *k-fold cross validation* yaitu 94% nilai presisi dan 95% nilai *recall* dan 90% nilai akurasi.

B. Algoritma Iterative Dechotomizer 3 (ID3)

Pada algoritma *Iterative Dechotomizer 3* (ID3) digunakan teknik uji *percentage split* metode latih *k-fold cross validation*, nilai *k* yang dipilih adalah 10. *Confussion Matrix* dari perhitungan *Iterative Dechotomizer 3* (ID3) dapat dilihat pada Tabel 5.

Tabel 5. *Confussion Matrix* Pada Algoritma *Iterative Dechotomizer 3* (ID3)

Kelas		Aktual teknik uji Percentage Split		Aktual teknik uji K-fold Cross Validation	
		YES (<i>positif lung cancer</i>)	NO (<i>Negatif Lung Cancer</i>)	YES (<i>positif lung cancer</i>)	NO (<i>Negatif Lung Cancer</i>)
Prediksi	YES (<i>positif lung cancer</i>)	48	6	262	8
	NO (<i>Negatif Lung Cancer</i>)	4	4	16	23

Dari Tabel 5, dapat dilihat bahwa algoritma *Iterative Dechotomizer 3* (ID3) dengan teknik *percentage split* memberikan hasil 48 kasus YES (*positif lung cancer*) diprediksi YES (*positif lung cancer*), 6 kasus YES (*positif lung cancer*) diprediksi NO (*negatif lung cancer*), 4 kasus NO (*negatif lung cancer*) diprediksi YES (*positif lung cancer*), dan 4 kasus NO (*negatif lung cancer*) diprediksi NO (*negatif lung cancer*). Selain itu, dari tabel 5 dapat dihitung nilai akurasi berdasarkan data yang berhasil diprediksi dengan tepat sebesar 87%. Hasil pengukuran *precision* untuk kelas YES (*positif lung cancer*) adalah sebesar 92% dan untuk kelas NO (*negatif lung cancer*) adalah sebesar 40%. Hasil pengukuran *recall* untuk kelas YES (*positif lung cancer*) yaitu 89% dan untuk kelas NO (*negatif lung cancer*) yaitu 50%.

Dari perhitungan ID3 menggunakan teknik uji *percentage split* diperoleh aturan linguistik untuk penyakit kanker paru-paru yaitu sebagai berikut:

Linguistic Rules:

IF ALLERGY $\leq$ 0.50 **AND** WHEEZING $\leq$ 0.50 **AND** PEER_PRESSURE $\leq$ 0.50 **AND** ANXIETY $\leq$ 0.50 **AND** SHORTNESS OF

BREATH $\leq$ 0.50 **AND** AGE $\leq$ 15.50 **THEN** NO.

IF ALLERGY $\leq$ 0.50 **AND** WHEEZING $\leq$ 0.50 **AND** PEER_PRESSURE $\leq$ 0.50 **AND** ANXIETY $\leq$ 0.50 **AND** SHORTNESS OF BREATH $\leq$ 0.50 **AND** AGE $\leq$ 15.50 **THEN** NO.

IF ALLERGY $\leq$ 0.50 **AND** WHEEZING $\leq$ 0.50 **AND** PEER_PRESSURE $\leq$ 0.50 **AND** ANXIETY $\leq$ 0.50 **AND** SHORTNESS OF BREATH $\leq$ 0.50 **AND** AGE $>$ 15.50 **THEN** YES.

IF ALLERGY $\leq$ 0.50 **AND** WHEEZING $\leq$ 0.50 **AND** PEER_PRESSURE $\leq$ 0.50 **AND** ANXIETY $\leq$ 0.50 **AND** SHORTNESS OF BREATH $>$ 0.50 **AND** AGE **THEN** NO.

IF ALLERGY $\leq$ 0.50 **AND** WHEEZING $\leq$ 0.50 **AND** PEER_PRESSURE $\leq$ 0.50 **AND** ANXIETY $>$ 0.50 **AND** CHRONIC DISEASE $\leq$ 0.50 **AND** AGE $\leq$ 18.50 **AND** ALCOHOL CONSUMING $\leq$ 0.50 **AND** AGE $\leq$ 9.5 **AND** GENDER $\leq$ 0.50 **THEN** YES.

IF ALLERGY $\leq$ 0.50 **AND** WHEEZING $\leq$ 0.50 **AND** PEER_PRESSURE $\leq$ 0.50 **AND** ANXIETY $>$ 0.50 **AND** CHRONIC DISEASE $\leq$ 0.50 **AND** AGE $\leq$ 18.50 **AND** ALCOHOL CONSUMING $\leq$ 0.50 **AND** AGE $\leq$ 9.5 **AND** GENDER $>$ 0.50 **THEN** NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND ANXIETY $>$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 AND AGE $\leq$ 18.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $>$ 9.5 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND ANXIETY $>$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 AND AGE $\leq$ 18.50 AND ALCOHOL CONSUMING $>$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND ANXIETY $>$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 AND AGE $>$ 18.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND ANXIETY $>$ 0.50 AND CHRONIC DISEASE $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $>$ 0.50 AND SWALLOWING DIFFICULTY $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND ANXIETY $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $>$ 0.50 AND SWALLOWING DIFFICULTY $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND ANXIETY $\leq$ 0.50 AND CHEST PAIN $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $>$ 0.50 AND SWALLOWING DIFFICULTY $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND ANXIETY $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $>$ 0.50 AND SWALLOWING DIFFICULTY $\leq$ 0.50 AND YELLOW_FINGERS $>$ 0.50 AND ANXIETY $\leq$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE $>$ 0.50 AND SWALLOWING DIFFICULTY $\leq$ 0.50 AND YELLOW_FINGERS $>$ 0.50 AND ANXIETY $>$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $\leq$ 0.50 AND PEER_PRESSURE AND FATIGUE $>$ 0.50 THEN YES

IF ALLERGY $\leq$ 0.50 AND WHEEZING $>$ 0.50 AND ANXIETY $\leq$ 0.50 AND GENDER $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $\leq$ 24.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $>$ 0.50 AND ANXIETY $\leq$ 0.50 AND GENDER $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $>$ 24.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $>$ 0.50 AND ANXIETY $\leq$ 0.50 AND GENDER $\leq$ 0.50 AND ALCOHOL CONSUMING $>$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $>$ 0.50 AND ANXIETY $\leq$ 0.50 AND GENDER $>$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND WHEEZING $>$ 0.50 AND ANXIETY $>$ 0.50 THEN YES.

IF ALLERGY $>$ 0.50 AND AGE $\leq$ 1.00 THEN NO.

IF ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $\leq$ 16.50 THEN NO.

IF ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $>$ 16.50 THEN YES.

IF ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $>$ 0.50 AND FATIGUE $\leq$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 AND AGE $\leq$ 20.00 THEN YES.

IF ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $>$ 0.50 AND FATIGUE $\leq$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 AND AGE $>$ 20 AND AGE $\leq$ 26.50 THEN NO.

IF ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $>$ 0.50 AND FATIGUE $\leq$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 AND AGE $>$ 20 AND AGE $>$ 26.50 THEN YES.

IF ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $>$ 0.50 AND FATIGUE $\leq$ 0.50 AND CHRONIC DISEASE $>$ 0.50 THEN YES.

ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $\leq$ 0.50 AND ALCOHOL CONSUMING $>$ 0.50 AND FATIGUE $>$ 0.50 THEN YES.

ALLERGY $>$ 0.50 AND AGE $>$ 1.00 AND YELLOW_FINGERS $>$ 0.50 THEN YES.

Hasil uji coba menggunakan teknik uji *k-fold cross validation* memberikan 262 kasus *YES (positif lung cancer)* diprediksi *YES (positif lung cancer)*, 8 kasus *YES (positif lung cancer)* diprediksi *NO (negatif lung cancer)*, 16 kasus *NO (negatif lung cancer)* diprediksi *YES (positif lung cancer)*, dan 23 kasus *NO (negatif lung cancer)* diprediksi *NO (negatif lung cancer)*. Selain itu, dari Tabel 5 dapat dihitung nilai akurasi diperoleh berdasarkan data yang berhasil diprediksi dengan tepat sebesar 92%. Hasil pengukuran *precision* untuk kelas *YES (positif lung cancer)* sebesar 94% dan untuk kelas *NO (negatif lung cancer)* sebesar 74%. Hasil pengukuran *recall* untuk kelas *YES (positif lung cancer)* yaitu 97% dan untuk kelas *NO (negatif lung cancer)* yaitu 59%.

Dari perhitungan ID3 menggunakan teknik uji *k-fold cross validation* diperoleh aturan linguistik untuk penyakit kanker paru-paru yaitu sebagai berikut:

Linguistic Rules:

IF ALLERGY $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND AGE $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND AGE $\leq$ 0.50 AND ALLERGY $\leq$ 0.50 THEN YES.

IF ALLERGY $>$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND AGE $\leq$ 0.50 AND ALLERGY $\leq$ 0.50 THEN NO.

IF ALLERGY $>$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND SHORTNESS OF BREATH $\leq$ 0.50 THEN YES.

IF ALLERGY $>$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND SHORTNESS OF BREATH $>$ 0.50 AND AGE $\leq$ 0.50 THEN YES.

IF ALLERGY $>$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 AND SHORTNESS OF BREATH $>$ 0.50 AND AGE $>$ 0.50 AND AGE $\geq$ 0.50 THEN YES.

IF ALLERGY $>$ 0.50 AND PEER_PRESSURE $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50

AND SHORTNESS OF BREATH $>$ 0.50 AND AGE $>$ 0.50 AND AGE $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $>$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL CONSUMING $\leq$ 0.50 AND AGE $>$ 0.50 AND YELLOW_FINGERS $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\leq$ 0.50 AND YELLOW_FINGERS $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\leq$ 0.50 AND YELLOW_FINGERS $>$ 0.50 AND CHORONIC DISEASE $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\leq$ 0.50 AND YELLOW_FINGERS $>$ 0.50 AND CHORONIC DISEASE $>$ 0.50 AND AGE $\leq$ 0.50 AND AGE $\leq$ 0.50 THEN YES.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\leq$ 0.50 AND YELLOW_FINGERS $>$ 0.50 AND CHORONIC DISEASE $>$ 0.50 AND AGE $\leq$ 0.50 AND AGE $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\leq$ 0.50 AND YELLOW_FINGERS $>$ 0.50 AND CHORONIC DISEASE $>$ 0.50 AND AGE $>$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $>$ 0.50 THEN NO

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\geq$ 0.50 AND AGE $\geq$ 0.50 THEN NO.

IF ALLERGY $\leq$ 0.50 AND ALCOHOL COMSUMING $\leq$ 0.50 AND CHEST PAIN $\leq$ 0.50 AND SHORTNESS OF BREATH $\geq$ 0.50 AND AGE $\leq$ 0.50 AND AGE $\leq$ 0.50 AND CHRONIC DISEASE $\leq$ 0.50 THEN YES.

IF ALLERGY ≤ 0.50 **AND** ALCOHOL COMSUMING ≤ 0.50 **AND** CHEST PAIN ≤ 0.50 **AND** SHORTNESS OF BREATH ≥ 0.50 **AND** AGE ≤ 0.50 **AND** AGE ≤ 0.50 **AND** CHRONIC DISEASE >0.50 **THEN** NO.

IF ALLERGY ≤ 0.50 **AND** ALCOHOL COMSUMING ≤ 0.50 **AND** CHEST PAIN ≤ 0.50 **AND** SHORTNESS OF BREATH ≥ 0.50 **AND** AGE ≤ 0.50 **AND** AGE >0.50 **THEN** NO

IF ALLERGY ≤ 0.50 **AND** ALCOHOL COMSUMING ≤ 0.50 **AND** CHEST PAIN ≤ 0.50 **AND** SHORTNESS OF BREATH ≤ 0.50 **AND** YELLOW_FINGERS ≤ 0.50 **THEN** YES

IF ALLERGY ≤ 0.50 **AND** ALCOHOL COMSUMING ≤ 0.50 **AND** CHEST PAIN ≤ 0.50 **AND** SHORTNESS OF BREATH ≤ 0.50 **AND** YELLOW_FINGERS >0.50 **THEN** NO.

Secara lebih lengkap, hasil *precision*, *recall*, dan *accuracy* algoritma ID3 dapat dilihat pada Tabel 6.

Tabel 6. Hasil Permodelan Algoritma *Iterative Dechotomizer 3* (ID3)

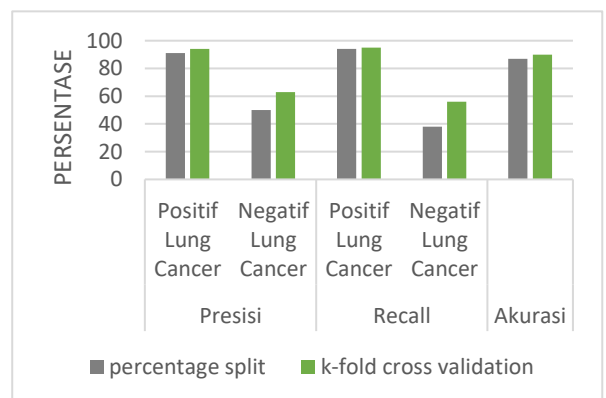
Target Class	Percentage split			K-fold cross validation		
	Presisi (%)	Recall (%)	Akurasi (%)	Presisi(%)	Recall (%)	Akurasi (%)
YES (positif lung cancer)	92	89	84	94	97	92
NO (Negatif Lung Cancer)	40	50		74	59	

Dari Tabel 6, diketahui bahwa teknik uji *k-fold cross validation* memiliki nilai *precision*, *recall* dan *accuracy* yang lebih baik dari teknik uji *percentage split*. Nilai presisi dari teknik uji *percentage split* adalah 92% sedangkan pada teknik uji *k-fold cross validation* 94%. Nilai *recall percentage split* adalah 89% dan nilai *recall k-fold cross validation* adalah 97%. Nilai *accuracy percentage split* sebesar 84% dan nilai *accuracy k-fold cross validation* sebesar 92%.

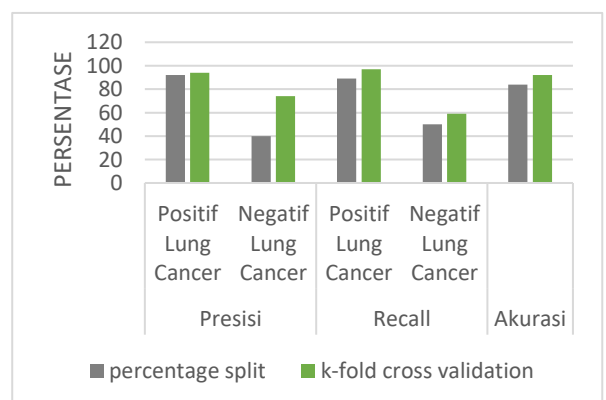
C. Perbandingan Hasil Kedua Algoritma

Hasil klasifikasi dua algoritma *naïve bayes* dan *iterative dechotomizer 3* (ID3), terlihat bahwa kedua metode tersebut bekerja dengan baik dalam melakukan klasifikasi kanker paru-paru pada dataset *Kaggle*. Perbandingan hasil pengukuran algoritma *naive bayes* dengan teknik uji *percentage split* dan *k-fold cross validation* pada Gambar 1. Dapat dilihat pada Gambar 1, algoritma *naive bayes* memiliki hasil nilai yang lebih baik pada teknik uji *k-fold cross validation* dibanding dengan teknik uji *percentage split*. Nilai akurasi pada teknik uji *k-fold cross validation* adalah 90% sedangkan pada teknik uji *percentage split* hanya 87%. Nilai presisi yang dihasilkan pada teknik uji *k-fold cross validation* adalah 94% sedangkan pada teknik uji *percentage split* hanya 91%.

Pada teknik uji *k-fold cross validation* diberikan nilai *recall* sebesar 95%, sedangkan pada teknik uji *percentage split* hanya 94%. Perbandingan hasil pengukuran algoritma *iterative dichotomizer 3* (ID3) dengan teknik uji *percentage split* dan *k-fold cross validation* pada Gambar 2 sebagai berikut.

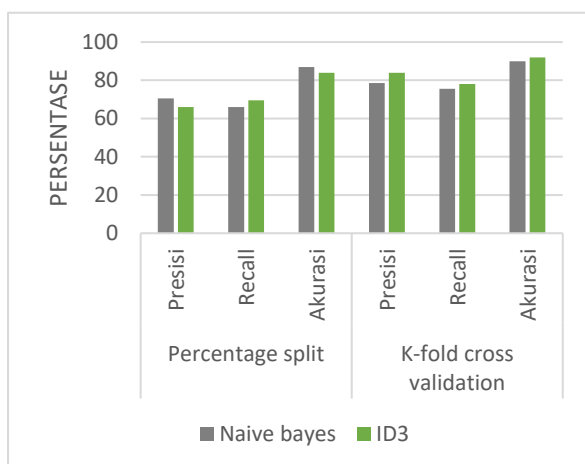


Gbr.1 Nilai-Nilai Akurasi, Presisi, *Recall* Naive Bayes



Gbr.2 Nilai-Nilai Akurasi, Presisi, dan *Recall* *Iterative Dechotomizer 3* (ID3)

Pada Gambar 2, dapat diketahui bahwa algoritma *Iterative Dichotomizer 3* (ID3) memperoleh hasil nilai uji yang lebih baik pada teknik *k-fold cross validation* dibanding dengan teknik uji *percentage split*. Nilai presisi pada teknik uji *k-fold cross validation* sebesar 94% sedangkan pada teknik uji *percentage split* hanya 92%. Nilai *recall* yang dihasilkan dari teknik uji *k-fold cross validation* sebesar 97% sedangkan pada teknik uji *percentage split* hanya 89%. Nilai akurasi yang diperoleh pada teknik uji *k-fold cross validation* sebesar 92% sedangkan pada teknik uji *percentage split* hanya 84%. Perbandingan rata-rata dengan teknik uji *percentage split* dan *k-fold cross validation* ditunjukkan pada Gambar 3.



Gbr.3 Perbandingan Rata-Rata Akurasi, Presisi, Recall Naive Bayes dan *Iterative Dichotomizer 3* (ID3)

Dari Gambar 3, dapat dilihat bahwa hasil perhitungan menggunakan teknik uji *k-fold cross validation* lebih baik dibanding dengan hasil teknik uji *percentage split*. Nilai akurasi pada teknik uji *percentage split* adalah 87% dan 84% untuk algoritma Naive Bayes dan ID3. Pada teknik uji *k-fold cross validation* diperoleh nilai akurasi sebesar 90% dan 92% untuk algoritma Naive Bayes dan ID3. Dari Gambar 3, dapat diketahui bahwa teknik uji *k-fold cross validation* lebih baik dari *percentage split*. Dari penelitian yang telah dilakukan dengan melihat Gambar 3 dapat diketahui bahwa ketika menggunakan teknik uji *percentage split*, algoritma naive Bayes memberikan hasil yang lebih baik dibandingkan algoritma ID3.

Algoritma ID3 memberikan hasil yang lebih baik dibandingkan dengan Naive Bayes ketika menggunakan teknik uji *k-fold cross validation*. Nilai akurasi dan *recall* yang diperoleh algoritma ID3 dengan teknik uji *k-fold cross validation* lebih baik dibandingkan dengan algoritma naive Bayes.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan menggunakan teknik uji *percentage split* diperoleh nilai terbaik pada algoritma naive Bayes untuk kelas YES (*positif lung cancer*) dengan nilai *accuracy* sebesar 87%, nilai *precision* 91% dan nilai *recall* sebesar 94%. Hasil uji teknik *k-fold cross validation* diperoleh nilai terbaik pada algoritma *iterative dichotomizer 3* (ID3) untuk kelas YES (*positif lung cancer*) yaitu sebesar 92% untuk nilai *accuracy*, 94% untuk nilai *precision*, dan 97% untuk nilai *recall*. Hasil yang diperoleh dari teknik uji *k-fold cross validation* memiliki nilai lebih tinggi dibandingkan dengan nilai dari teknik uji *percentage split*. Kesimpulan dari penelitian ini, diperoleh algoritma terbaik yaitu ID3 dengan teknik uji *k-fold cross validation* pada dataset kanker paru-paru.

REFERENSI

- [1] Juwita, N. Amalita, dan M. P. Dewi, "Faktor-Faktor Risiko yang Mempengaruhi Kanker Paru-Paru dengan Menggunakan Analisis Regresi Logistik," *J. Math. UNP*, vol. 4, no. 1, hal. 38–42, 2021.
- [2] A. Antoni, A. Pujiarti, A. Putri, A. Prayoga, D. Yolanda, dan E. Sari, "Penyuluhan Kesehatan: Mengenal Kanker Paru di SMPN 3 Padangsidempuan," vol. 3, no. 3, hal. 175–177, 2022.
- [3] Nova, "Asuhan Keperawatan Pada Klien Efusi Pleura Dengan Masalah Keperawatan Ketidakefektifan Pola Nafas Di Ruang Teratai Rumah Sakit Umum Daerah Bangil Pasuruan," *StikesInsanCendekiaMedika*, vol. 8, no. 5, hal. 55, 2019.
- [4] L. Sari, A. Romadloni, dan R. Listyaningrum, "Penerapan Data Mining dalam Analisis Prediksi Kanker

- Paru Menggunakan Algoritma Random Forest,” *Infotekmesin*, vol. 14, no. 01, hal. 155–162, 2023.
- [5] E. Amalia, D. P. Sari, Y. P. Susani, dan B. V. Fidia, “Pelatihan *Psychological Self Care* pada Petugas Kesehatan yang Melakukan Perawatan pada Pasien COVID-19 di Rumah Sakit Umum Daerah Provinsi NTB,” *J. Pengabd. Magister Pendidik. IPA*, vol. 3, no. 2, hal. 0–5, 2020.
- [6] W. Nugraha, “Prediksi Penyakit Jantung Cardiovascular Menggunakan Model Algoritma Klasifikasi,” *J. Manag. dan Inform.*, vol. 9, no. 2, hal. 78–84, 2021.
- [7] Alvina Felicia Watratan, Arwini Puspita. B, dan Dikwan Moeis, “Implementasi Algoritma Naive Bayes Untuk Memprediksi Tingkat Penyebaran Covid-19 Di Indonesia,” *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 1, hal. 7–14, 2020.
- [8] A. Desiani, R. Primartha, M. Arhami, dan O. Orsalan, “Naive Bayes classifier for infant weight prediction of hypertension mother,” *J. Phys. Conf. Ser.*, vol. 1282, no. 1, hal. 1–9, 2019.
- [9] Riska Agustin, Vera Maya Santi, dan Bagus Sumargo, “Metode Naive Bayes dalam Mendeteksi Sel Kanker Payudara,” *J. Stat. dan Apl.*, vol. 3, no. 1, hal. 30–38, 2019.
- [10] H. Oktavianto dan R. P. Handri, “Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes,” *INFORMAL Informatics J.*, vol. 4, no. 3, hal. 117, 2020.
- [11] A. Ridwan, “Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, hal. 15–21, 2020.
- [12] A. Byna dan M. Basit, “Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naive Bayes,” *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, hal. 407–411, 2020.
- [13] A. K. Nugroho dan D. Iskandar, “Algoritma Iterative Dichotomizer 3 (ID3) Pengambilan Keputusan Decision Making Using Iterative Dichotomizer 3 Algorithm,” *Din. Rekayasa*, vol. 11, no. 2, hal. 43–47, 2015.
- [14] F. Hafidh, M. Y. Kurniawan, dan R. I. Yazidah Anwar, “Identifikasi Ketunaan Anak Berkebutuhan Khusus dengan Algoritma Iterative Dichotomiser 3 (ID3),” *J. Buana Inform.*, vol. 12, no. 2, hal. 78–87, 2021.
- [15] H. Hikmatulloh, A. Rahmawati, D. Wintana, dan D. A. Ambarsari, “Penerapan Algoritma Iterative Dichotomiser Three (Id3) Dalam Mendiagnosa Kesehatan Kehamilan,” *Klik - Kumpul. J. Ilmu Komput.*, vol. 6, no. 2, hal. 116, 2019.
- [16] K. Ma'mur, “Analisis Penerapan Algoritma ID3 dalam Mendiagnosis Kesuburan Pria,” *J. Inform. Univ. Pamulang*, vol. 4, no. 2, hal. 35, 2019.
- [17] D. Balai Kesehatan Kementerian Perindustrian Jakarta, Avia Enggar Tyasti, D. Ispriyanti, dan A. Hoyyi, “Algoritma Iterative Dichotomiser 3 (ID3) untuk Mengidentifikasi Data Rekam Medis (Studi Kasus Penyakit Diabetes Mellitus),” *J. gaussian*, vol. 4, no. 2, hal. 237–246, 2015.
- [18] I. Print, A. F. Pulungan, dan D. Selvida, “Kombinasi Metode Sampling pada Pengklasifikasian Data Tidak Seimbang Menggunakan Algoritma Support Vector Machine (SVM),” *InfoTekJar J. Nas. Inform. dan Teknol. Jar.*, vol. 6, no. 2, hal. 276–282, 2022.
- [19] M. Furqan, S. Sriani, dan S. M. Sari, “Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia,” *Techno.Com*, vol. 21, no. 1, hal. 51–60, 2022.
- [20] Y. T. U. Heni Sulistiani, “Penerapan Algoritma Klasifikasi Sebagai Pendukung Keputusan Pemberian Beasiswa Mahasiswa,” *Snti*, no. October 2018, hal. 300–305, 2018.
- [21] H. Wahyu Sulisty, H. Oktavianto Universitas Muhammadiyah Jember, dan J. Karimata, “Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Random Forest,” *J. Inform.*, vol. 8, no. 2, hal. 45–51, 2020.
- [22] S. P. Nabila, N. Ulinuha, dan A. Yusuf, “Model Prediksi Kelulusan Tepat

- Waktu Dengan Metode Fuzzy C-Means Dan K-Nearest Neighbors Menggunakan Data Registrasi Mahasiswa,” *Netw. Eng. Res. Oper.*, vol. 6, no. 1, hal. 39, 2021.
- [23] D. F. Zhafira, B. Rahayudi, dan I. Indriati, “Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube,” *J. Sist. Informasi, Teknol. Informasi, dan Edukasi Sist. Inf.*, vol. 2, no. 1, hal. 55–63, 2021.
- [24] Normah, B. Rifai, S. Vambudi, dan R. Maulana, “Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis SMOTE,” *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, hal. 174–180, 2022.
- [25] D. Putra dan A. Wibowo, “Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes,” *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, no.147, hal. 84–92, 2020.
- [26] D. Normawati dan S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, hal. 697–711, 2021.
- [27] M. F. Rifai, H. Jatnika, dan B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” *Petir*, vol. 12, no. 2, hal. 131–144, 2019.
- [28] H. Hozairi, A. Anwari, dan S. Alim, “Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes,” *Netw. Eng. Res. Oper.*, vol. 6, no. 2, hal. 133, 2021.
- [29] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, “Analisis Sentimen Masyarakat terhadap Hasil *Quick Count* Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier,” *Indones. J. Appl. Stat.*, vol. 2, no. 1, pp. 34, 2019.
- [30] R. K. Dinata, S. Safwandi, N. Hasdyna, dan N. Azizah, “Analisis K-Means Clustering pada Data Sepeda Motor,” *INFORMAL Informatics J.*, vol. 5, no. 1, hal. 10, 2020.