

Komparasi Kinerja Algoritma Machine Learning Untuk Deteksi Penyakit Infeksi Saluran Pernapasan

Muh. Jamil*¹, Fahrul Rozi², Yudhi Fajar Saputra³

^{1,3}Ilmu Komputer, Fakultas Teknik dan Informatika, Universitas Widya Gama Mahakam Samarinda

²Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri
e-mail: *¹jamil@uwgm.ac.id, ²rozysmd@gmail.com, ³fajaryudhi@uwgm.ac.id

Abstrak

Infeksi Saluran Pernafasan Akut (ISPA) menjadi salah satu penyakit infeksi penyebab utama kematian di dunia. Hampir empat juta orang meninggal karena ISPA setiap tahunnya, 98% disebabkan oleh infeksi saluran pernapasan bawah. Angka kematian sangat tinggi pada bayi, anak-anak dan orang tua, terutama di negara-negara dengan pendapatan per kapita rendah dan menengah. ISPA di Indonesia menempati urutan pertama penyebab kematian pada anak dan dewasa. ISPA juga berada pada daftar 10 penyakit terbanyak. Survei yang dilakukan menunjukkan bahwa ISPA merupakan salah satu penyebab kematian anak terbesar di Indonesia dengan persentase 22,30. Tujuan penelitian adalah untuk membandingkan kinerja beberapa algoritma machine learning untuk mendeteksi penyakit ISPA. Proses klasifikasi yang dilakukan berdasarkan beberapa gejala dari pasien yang didapatkan dari data rekam medis beberapa pasien di Puskesmas Lojanan Kabupaten Kutai Kartanegara sebagai sumber utama dataset pada penelitian ini. Hasil uji klasifikasi untuk masing-masing kelas dengan algoritma yang digunakan memunculkan hasil terendah pada algoritma Naive Bayes yaitu 98% sementara kinerja terbaik dihasilkan oleh algoritma KNN, ANN, Random Forest, dan SVM dengan nilai akurasi 100%. Hasil ini sudah mampu mengungguli hasil kinerja dari penelitian sebelumnya dengan kasus dan jumlah kelas yang sama.

Kata kunci— Ispa, Klasifikasi, Machine learning

1. PENDAHULUAN

Kecerdasan buatan adalah bidang dalam ilmu komputer yang ditujukan untuk menciptakan sebuah kecerdasan pada mesin komputer layaknya kecerdasan yang dimiliki oleh manusia. Sehingga kecerdasan buatan banyak digunakan dalam memecahkan permasalahan dalam bidang pekerjaan manusia, salah satunya adalah pada bidang kesehatan dan medis [1]. *Machine learning* adalah salah satu cabang keilmuan dari kecerdasan buatan yang mengadopsi prinsip ilmu komputer dan statistik untuk menciptakan model yang merefleksikan pola dari sebuah data masukan. Dalam proses pembuatan model tersebut dapat menggunakan berbagai macam algoritma cerdas yang dapat disesuaikan dengan kebutuhan [2]. Semisal jika dengan kasus klasifikasi dapat menggunakan algoritma klasifikasi seperti *Decision Tree*, *Naive Bayes*, *Random Forest* dan *Support Vectore Machine (SVM)* dan masih banyak lagi [3]. Atau jika dengan kasus *clustering* dapat menggunakan algoritma *K-Means*, *DBSCAN* [4] dan juga *K-Medoids* [5] dan masih banyak lagi.

Infeksi saluran pernapasan (ISPA) adalah penyakit akut dan bahkan dapat menyebabkan kematian. Penyakit ini dapat ditandai dengan gejala yang bervariasi diawali dengan peningkatan

suhu tubuh dan disertai dengan salah satu atau lebih gejala lain seperti tenggorokan sakit, nyeri tenggorokan saat menelan, batuk pilek yang dapat disertai dengan dahak ataupun tidak [6]. Menurut data WHO pada tahun 2016, angka kejadian penyakit ISPA di Indonesia adalah 151 juta jiwa pertahun. Menurut data kemenkes pada tahun 2017 ISPA menempati urutan 10 besar rawat inap di rumah sakit dan masuk 4 dari 10 besar penyakit di wilayah puskesmas [7]. ISPA adalah jenis penyakit yang cukup kompleks yang disebabkan oleh beberapa pemicu, bisa disebabkan oleh virus seperti *Rotavirus*, virus *Influenza* serta bakteri, penyebab lain adalah pengaruh atmosfer seperti pencemaran lingkungan [8].

Penelitian mengenai penggunaan teknik data mining dalam proses identifikasi penyakit ISPA sudah banyak dilakukan oleh peneliti sebelumnya. Pada penelitian [9] metode *Fuzzy* dan *KNN* digunakan untuk klasifikasi penyakit ISPA ringan dan berat. Kemudian pada penelitian [10] dan [11] menggunakan algoritma *C4.5* untuk klasifikasi penyakit ISPA. Selain itu penelitian [12] yang melakukan perbandingan kinerja beberapa algoritma untuk klasifikasi penyakit ISPA dengan hasil terbaik pada algoritma *SVM* dengan akurasi sebesar 99%. Perbandingan lain yang pernah dilakukan oleh penelitian [13] menggunakan algoritma klasifikasi yang memberikan hasil bahwa algoritma *SVM* memiliki kinerja terbaik dalam melakukan proses klasifikasi penyakit ISPA dengan tingkat akurasi sebesar 100%.

Berdasarkan hasil kinerja algoritma di penelitian terdahulu yang melakukan perbandingan beberapa algoritma klasifikasi untuk mendeteksi penyakit ISPA. Maka, penelitian ini akan melakukan hal yang sama yaitu membandingkan kinerja algoritma klasifikasi *Naive Bayes* dan *SVM*, *ANN*, *KNN* dengan penambahan 1 algoritma klasifikasi yang berbeda dari dua penelitian sebelumnya yaitu algoritma *Random forest* untuk klasifikasi penyakit ISPA. Dataset yang digunakan merupakan data pasien ISPA dari Puskesmas Loajanan Kabupaten Kutai Kartanegara. Dengan menggunakan beberapa matriks performansi maka akan diukur kinerja dari masing-masing algoritma yang digunakan.

2. METODE PENELITIAN

2.1. Pengumpulan dan preprosesing dataset

Dataset yang digunakan adalah dataset riwayat penyakit pasien dari sebuah puskesmas yang berada di Kecamatan Loa Janan Kabupaten Kutai Kartanegara, Provinsi Kalimantan Timur. Dataset inilah yang kemudian akan diujicobakan terhadap algoritma yang akan dibandingkan. Dataset yang ada kemudian akan dipilah seperti golongan penyakit serta beberapa atribut yang memiliki keterhubungan dengan diagnosa ISPA. Sumber data yang digunakan adalah data rekam medis pasien dengan rentang waktu dari tahun 2018 sampai 2021. Data yang ada kemudian dipindahkan ke lembar kerja excel untuk dilakukan *preprocessing* dan disesuaikan angka dan nilainya, kemudian dirubah kedalam bentuk file *csv* agar dapat diolah oleh tool dan bahasa pemrograman yang digunakan.

2.2. Metode algoritma klasifikasi

Terdapat beberapa algoritma klasifikasi yang akan dibandingkan dalam penelitian ini dengan menggunakan matriks performansi yang sama. Empat algoritma yang terinspirasi dari penelitian [13] dan penelitian [12] sebelumnya, yaitu *Naive Bayes*, *KNN*, *ANN* dan *SVM* dan dengan penambahan satu algoritma pada penelitian ini yaitu *Random Forest* dengan jumlah kelas yang sama yaitu positif ISPA dan negative ISPA. Algoritma *Random Forest* dipilih karena algoritma ini memiliki kinerja yang sangat baik pada beberapa penelitian dengan kasus klasifikasi 2 kelas ataupun lebih seperti pada penelitian [14,15]. Penelitian ini menggunakan *scikit learn* [16]

sebagai library yang dapat mempermudah dalam mengimplementasikan algoritma yang akan diuji cobakan dalam proses klasifikasi. Detail dari setiap algoritma yang akan digunakan adalah sebagai berikut:

a. *K-Nearesr Neighbors (KNN)*

Adalah algoritma klasifikasi yang termasuk ke dalam algoritma *supervised learning*. *KNN* juga merupakan salah satu contoh dari *instance-based learning*, salah satu langkah terpenting dalam penggunaan algoritma *KNN* adalah menentukan nilai *k* pada data *training*. Semisal jika nilai $k = 5$ maka akan diambil 5 kandidat data dari data *training* yang memiliki nilai paling dekat dengan data yang akan diprediksi [17].

b. *Naive Bayes*

Asumsi dalam algoritma *Naive Bayes* adalah bahwa semua fitur pada data *training* bersifat independen satu dengan yang lain. *Naive Bayes* menggunakan prinsip statistik dan matematika untuk memprediksi probabilitas keanggotaan baru, setelah membandingkan nilai probabilitas dari setiap kelas yang mungkin, algoritma ini akan memilih nilai probabilitas tertinggi sebagai prediksi [18,19].

c. *Artificial Neural Network (ANN)*

Algoritma ini terinspirasi dari bentuk jaringan saraf manusia. Yang di mana sebuah sel saraf atau *neuron* terdiri dari beberapa *layer* yaitu *input layer*, *hidden layer* dan *output layer* yang dibekali dengan fungsi aktivasi untuk menentukan hubungan *non-linear* pada data serta adanya *backpropagation* dalam melakukan optimalisasi bobot untuk mengurangi nilai kesalahan dalam prediksi [20].

d. *Support Vector Machine (SVM)*

Konsep kerja dari *SVM* adalah untuk menemukan *hyperplane* atau batas keputusan (*decision boundary*) yang seoptimal mungkin. Hal ini bertujuan untuk memisahkan kelompok data menjadi dua kelas secara *linear*, *hyperplan* akan memiliki sebuah garis atau margin yang akan selalu dimaksimalkan jaraknya untuk memisahkan data pada setiap kelas [21].

e. *Random Forest*

Algoritma *Random Forest* menggunakan metode pemisahan *biner rekursif* untuk menuju node akhir pada struktur pohon keputusan. Kinerja algoritma ini dikenal baik dan efisien dengan kondisi jumlah data *training* yang cukup besar. Algoritma ini mampu meminimalkan nilai *error* sehingga model mampu menghasilkan nilai akurasi yang seoptimal mungkin [22].

2.3. *Hyper parameter tuning*

Fungsi dari pemberian *hyper parameter* adalah untuk memberikan nilai optimum pada model *machine learning* yang digunakan [23]. Metode *tuning* yang akan digunakan adalah *GridSearchCV* [24] yang disediakan oleh library *scikit learn*. Menurut [25] metode *GridSearchCV* memiliki kemampuan dalam mencari optimasi *hyper parameter* terbaik pada model *machine learning* yang digunakan. Pada penelitian sebelumnya pun seperti [26] menghasilkan nilai akurasi yang lebih baik dengan menggunakan metode *tuning GridSearchCV* dibandingkan dengan malakukan *hyper parameter tuning* secara manual. Untuk dapat melakukan metode *tuning* dengan *GridSearchCV* maka perlu diberikan beberapa kandidat *hyper parameter*. Kandidat inilah yang kemudian akan diuji satu persatu oleh library *GridSearchCV* hingga menghasilkan sebuah keluaran *hyper parameter* terbaik berdasarkan dataset dan metode algoritma yang digunakan. Pada tabel 1 menunjukkan kandidat *hyper parameter* yang diberikan pada setiap algoritma.

Tabel 1. kandidat *hyper parameter* untuk setiap algoritma

No.	Algoritma	Hyper Parameter
1	<i>KNN</i>	<i>n_neighbors = ranges 2-10</i> <i>algorithm = auto,ball_tree, kd_tree, brute</i> <i>weights = uniform,distance</i>
2	<i>ANN</i>	<i>hidden_layer_sizes = (5,2)</i> <i>activation = identity, logistic, tanh, relu</i> <i>solver = lbfgs, sgd, adam</i>
3	<i>Naive Bayes</i>	<i>random_state = ranges 1-10</i> <i>var_smoothing = 1,0, 0,1, 0,01, 0,001,</i> <i>0,0001, 1e-05, 1e-06, 1e-07, 1e-08, 1e-09</i>
5	<i>Random Forest</i>	<i>criterion = gini, entropy</i> <i>max_depth = ranges 1-10</i> <i>random_state = ranges 1-10</i>
6	<i>SVM</i>	<i>kernel = linear, poly, rbf, sigmoid</i>

2.4. Metode training dan testing

Metode *training* dan *testing* yang digunakan akan menggunakan dua kondisi. Yaitu dengan *train test split* dengan komposisi data 80% data *training* dan 20% data *testing* serta kondisi kedua yaitu dengan *10 Fold-Cross Validation*. data *training* adalah sekumpulan data berlabel yang diberikan sebagai input dan digunakan oleh model algoritma untuk mempelajari pola dan karakteristik data, sementara data *testing* adalah sekumpulan data yang memiliki fitur yang sama dengan data *training* namun dengan jumlah yang lebih sedikit. Data *testing* digunakan untuk memvalidasi kinerja algoritma yang digunakan [27]. Sementara *10 Fold-Cross Validation* bekerja dengan cara membagi data menjadi beberapa segmen *train* dan *testing* sehingga setiap data mendapatkan kesempatan yang sama menjadi data *testing*. Teknik ini banyak digunakan oleh peneliti untuk mengurangi bias pada hasil klasifikasi [28].

2.5. Metode validasi performa algoritma

Setelah masing-masing algoritma diuji dengan 2 model *training* dan *testing*. Maka dari prediksi yang dihasilkan akan divalidasi performanya dengan menggunakan parameter *precision*, *recall*, *f1-score* dan nilai *accuracy* yang dihasilkan.

a. Precision

Merupakan metrik performansi untuk mengevaluasi model yang dibuat dan menggambarkan prediksi dari kondisi sebenarnya dengan prediksi yang dihasilkan oleh algoritma yang digunakan [29]. Formula untuk precision ditunjukkan pada persamaan 1.

$$\text{pre} = \frac{\text{TP}}{\text{FP} + \text{TP}} \quad (1)$$

b. Accuracy

Adalah perbandingan antara hasil prediksi yang benar dari keseluruhan data yang telah diuji. Formula dari *accuracy* ditunjukkan pada persamaan 2.

$$\text{acc} = \frac{\text{TN}}{\text{FN} + \text{FP} + \text{TN} + \text{TP}} \quad (2)$$

c. Recall

Adalah ukuran sensitivitas dari data untuk setiap kelas yang diprediksi benar terhadap data yang salah diprediksi. Formula dari recall ditunjukkan pada persamaan 3.

$$\text{rec} = \frac{\text{TP}}{\text{FN} + \text{TP}} \quad (3)$$

d. *F1-Score*

Adalah perbandingan rata-rata *precision* dan *recall* yang dibobotkan. Formula dari F1-Score ditunjukkan pada persamaan 4.

$$F1 - Score = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (4)$$

3. HASIL DAN PEMBAHASAN

3.1 *Hasil pengumpulan dan preprocessing dataset*

Data yang digunakan berjumlah 253 *instance* dan terdiri dari 16 fitur. 15 fitur yang digunakan sebagai variable X adalah (kelompok umur, jenis kelamin, tekanan darah, tinggi badan, berat badan, denyut nadi, suhu badan, *respiration rate*, kondisi hidung, batuk, demam, pilek, sakit tenggorokan, sesak nafas, hidung tersumbat) 1 fitur sebagai variable Y (diagnosa). Komposisi dataset yang digunakan adalah 116 *instance* dengan diagnosa ispa dan 137 *instance* dengan diagnosa non ispa. Dataset yang telah dikumpulkan kemudian diubah kedalam bentuk *csv* agar dapat diolah menggunakan *tool Jupyter Notebook* dengan bahasa pemrograman *python*. Sampel dataset yang telah diproses ditunjukkan pada Tabel 2.

Tabel 2. Hasil *preprocessing dataset*

No	Kelompok Umur	Golongan Tekanan	Suhu Badan	...	Pilek	Sakit Tenggorokan	Diagnosis
1	3	2	37	...	1	1	1
2	1	2	36	...	1	1	1
3	3	2	37	...	1	1	1
4	2	2	37	...	1	1	1
5	2	3	38	...	1	1	1
...
249	2	1	36	...	0	0	0
250	2	1	36	...	0	0	0
251	2	1	36	...	0	0	0
252	2	2	36	...	0	0	0
253	2	1	37	...	0	0	0

Pada Tabel 2 menampilkan bentuk *dataset* yang akan digunakan. Nilai pada fitur telah diubah ke dalam bentuk angka namun dengan arti yang sama, hal ini dimaksudkan agar algoritma yang digunakan dapat membaca nilai pada kolom tersebut, semisal untuk fitur batuk, jika nilainya adalah 1 berarti pasien tersebut mengalami batuk, dan nilai 0 adalah kondisi pasien tidak mengalami gejala batuk.

3.2 *Hasil ujicoba klasifikasi dengan teknik train test split*

Pengujian yang dilakukan adalah dengan menerapkan beberapa algoritma *machine learning*. Pengujian pertama dilakukan dengan teknik *train test split* dengan komposisi 80% data *training* dan 20% data *test*. Setelah dilakukan *parameter tuning* dengan menggunakan metode *GridSearchCV* maka sudah didapatkan *parameter tuning* terbaik pada masing-masing algoritma yang digunakan. Hasil *parameter tuning* ditunjukkan pada Tabel 3.

Tabel 3. Hasil *parameter tuning*.

No	Algoritma	Hyper parameter
1	<i>KNN</i>	<i>n_neighbors = 2</i> <i>algorithm = auto</i> <i>weights = distance</i>

2	ANN	<i>hidden_layer_sizes</i> = (5,2) <i>activation</i> = <i>relu</i> <i>solver</i> = <i>lbfgs</i> <i>random_state</i> = 6
3	Naïve Bayes	<i>var_smoothing</i> = 1.0
5	Random Forest	<i>criterion</i> = <i>gini</i> <i>max_depth</i> = 6 <i>random_state</i> = 9
6	SVM	<i>kernel</i> = <i>rbf</i>

Dengan penggunaan *hyper parameter* pada Tabel 3 maka hasil kinerja yang didapatkan untuk masing-masing algoritma disajikan pada Tabel 4.

Tabel 4. Hasil pengujian kinerja algoritma

No	Algoritma	Precision		Recall		F1-Acore		Accuracy
		ispa	non ispa	ispa	non ispa	ispa	non ispa	
1	KNN	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	ANN	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3	Naïve Bayes	1,00	0,97	0,96	1,00	0,98	0,98	0,98
5	Random Forest	1,00	1,00	1,00	1,00	1,00	1,00	1,00
6	SVM	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Tabel 4 menampilkan kinerja dari setiap algoritma yang digunakan dalam melakukan klasifikasi untuk kelas Ispa dan Non Ispa. Performa terbaik ditunjukkan oleh algoritma SVM dan KNN, ANN dan Random Forest dengan akurasi sebesar 1,0 dan 1 algoritma yang memiliki akurasi sebesar 98% adalah Naïve Bayes.

3.3 Hasil uji klasifikasi dengan teknik 10 K-Fold Validation

Pengujian kedua dilakukan dengan menggunakan *10 K-Fold Validation*. Cara kerja dari teknik ini adalah akan membagi keseluruhan data menjadi 10 bagian dan kemudian data yang telah dibagi akan diuji persetiap bagiannya. Namun sebelum dilakukan pengujian dengan *10 K-fold validation* maka terlebih dahulu dilakukan *hyper parameter tuning* menggunakan metode *GridSearchCV* dengan hasil rekomendasi yang dapat dilihat pada Tabel 5.

Tabel 5. Rekomendasi *hyper parameter* yang didapatkan dengan metode *GridSearchCV*. Dengan teknik *10 K-Fold Validation*

No	Algoritma	Hyper parameter
1	KNN	<i>n_neighbors</i> = 6 <i>algorithm</i> = <i>auto</i> <i>weights</i> = <i>distance</i>
2	ANN	<i>hidden_layer_sizes</i> = (5,2) <i>activation</i> = <i>identity</i> <i>solver</i> = <i>lbfgs</i> <i>random_state</i> = 2
3	Naïve Bayes	<i>var_smoothing</i> = <i>1e-05</i>
5	Random Forest	<i>criterion</i> = <i>gini</i> <i>max_depth</i> = 3 <i>random_state</i> = 4
6	SVM	<i>kernel</i> = <i>linear</i>

Berdasarkan hasil pada Tabel 5 terdapat beberapa nilai *hyper parameter* yang berbeda sesuai dengan rekomendasi *GridSearchCV* dengan menggunakan *10 K-Fold Validation* dibandingkan dengan menggunakan teknik *train test split*. Hasil kinerja pada setiap algoritma disajikan pada Tabel 6.

Tabel 6. Hasil pengujian kinerja algoritma dengan teknik *10 K-Fold Validation*.

No	Algoritma	Precision		Recall		F1-Acore		Accuracy
		ispa	non ispa	ispa	non ispa	ispa	non ispa	
1	<i>KNN</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00
2	<i>ANN</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00
3	<i>Naïve Bayes</i>	0,98	0,99	0,98	0,99	0,98	0,99	0,98
5	<i>Random Forest</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00
6	<i>SVM</i>	1,00	1,00	1,00	1,00	1,00	1,00	1,00

Tabel 6 menampilkan performa algoritma yang digunakan dengan menggunakan *10 K-fold validation*. Hasil klasifikasi untuk masing-masing kelas dengan algoritma yang digunakan memunculkan hasil terbaik pada *Random Forest, KNN, ANN, SVM* dengan nilai akurasi 1,00, sementara algoritma *Naïve Bayes* mendapatkan akurasi sebesar 98%. Jika dibandingkan dengan teknik *train test split Naive Bayes* mengalami kenaikan kinerja sebesar 1% pada beberapa matriks performansi yang digunakan dengan menggunakan *10 K-fold validation*.

Jika melihat hasil penelitian sebelumnya seperti pada penelitian Ginantra NLWSR dkk [13]. dengan judul “*Machine learning approach for acute Respiratory infections (ISPA) prediction: Case study Indonesia*”. Menggunakan metode *KNN, SVM, Naïve Bayes*, dan *Neural Network* telah mendapatkan nilai akurasi yang baik dengan angka terendah adalah 93% pada metode *ANN* dan hasil terbaik pada *SVM* dengan akurasi sebesar 100%, meskipun demikian penelitian lain dengan pembahasan yang sama yaitu [12] dengan judul “*Komparasi Algoritma Untuk Klasifikasi Penyakit ISPA (Infeksi Saluran Pernapasan Akut)*” mendapatkan akurasi terendah pada algoritma *KNN* dengan nilai akurasi sebesar 94% dan nilai terbaik dihasilkan oleh algoritma *SVM* dengan akurasi sebesar 99%. Detail hasil perbandingan dari penelitian ini dengan penelitian sebelumnya ditunjukkan pada Tabel 7.

Tabel 7. Perbandingan hasil kinerja penelitian kami dengan penelitian terdahulu.

No	Penelitian	Jumlah Instance	Jumlah Fitur	Metode Algoritma	Accuracy
1	Ginantra NLWSR dkk [13]	150	15	<i>Naive Bayes</i>	0,94
				<i>KNN</i>	0,94
				<i>SVM</i>	1,00
				<i>ANN</i>	0,93
2	M. Napih dkk [12]	201	22	<i>Naive Bayes</i>	0,98
				<i>KNN</i>	0,94
				<i>SVM</i>	0,99
3	Penelitian yang dikerjakan	252	16	<i>Naive Bayes</i>	0,98
				<i>KNN</i>	1,00
				<i>SVM</i>	1,00
				<i>ANN</i>	1,00
				<i>Random Forest</i>	1,00

4. KESIMPULAN

Berdasarkan hasil perbandingan penelitian ini dengan penelitian sebelumnya yang disajikan pada Tabel 7. Menunjukkan bahwa terdapat beberapa perbedaan kinerja yang mana teknik yang digunakan mampu mengungguli penggunaan algoritma *Naive Bayes, KNN* dan *ANN* pada penelitian [12, 13], namun dapat diamati juga bahwa algoritma *SVM* di penelitian ini dan di

penelitian sebelumnya memiliki hasil kinerja terbaik. Penambahan algoritma *Random Forest* yang diuji coba juga menunjukkan hasil yang sangat optimal dengan nilai akurasi sebesar 100%.

5. SARAN

Terdapat beberapa saran untuk penelitian selanjutnya dalam melengkapi kekurangan yang telah dikerjakan pada penelitian ini. Di antaranya adalah:

1. Penelitian selanjutnya diharapkan mampu menambahkan jumlah kelas yang digunakan agar deteksi penyakit ISPA bisa menjadi lebih sempurna semisal penambahan kelas non ISPA, ISPA ringan, ISPA berat.
2. Penelitian selanjutnya diharapkan dapat menambah jumlah dataset. Tidak hanya berasal dari satu sumber puskesmas saja, namun bisa menambahkan data dari puskesmas lain ataupun dari rumah sakit.

UCAPAN TERIMA KASIH

Ucapan terima kasih yang sebesar besarnya diucapkan kepada pihak Puskesmas Loajanan Kabupaten Kutai Kartanegara yang telah membantu dalam menghimpun dataset yang digunakan dalam menyelesaikan penelitian ini.

DAFTAR PUSTAKA

- [1] A. Roihan, P. Abas Sunarya, and A. S. Rafika, "IJCIT (Indonesian Journal on Computer and Information Technology) Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper," 2019.
- [2] N. Giarsyani, A. F. Hidayatullah, and R. Rahmadi, "Komparasi Algoritma Machine Learning Dan Deep Learning Untuk Named Entity Recognition : Studi Kasus Data Kebencanaan," *Jurnal Informatika dan Rekayasa Elektronik*, vol. 3, no. 1, pp. 48–57, Aug. 2020, Accessed: Aug. 13, 2024. [Online]. Available: <https://e-journal.stmiklombok.ac.id/index.php/jire/article/view/222>
- [3] N. B. Putri and A. W. Wijayanto, "Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing," *Komputika : Jurnal Sistem Komputer*, vol. 11, no. 1, pp. 59–66, Jan. 2022, doi: 10.34010/KOMPUTIKA.V11I1.4350.
- [4] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *SITEKIN: Jurnal Sains, Teknologi dan Industri*, vol. 18, no. 2, pp. 206–211, Jun. 2021, doi: 10.24014/SITEKIN.V18I2.12469.
- [5] L. F. olivia, D. A. Juliantho, and B. Hendrik, "Komparasi Algoritma K-Means Dan K-Medoids Dalam Clustering Penyebaran Kasus Covid 19," *Journal of Information System and Education Development*, vol. 1, no. 2, pp. 30–32, Sep. 2023, doi: 10.31849/digitalzone.v12i1.6572ICCS.
- [6] D. Talok Kecamatan Kresek *et al.*, "Hubungan Faktor Lingkungan dan Perilaku Terhadap Kejadian Infeksi Saluran Pernafasan Akut (ISPA) di Desa Talok Kecamatan Kresek," *Jurnal Kedokteran YARSI*, vol. 27, no. 2, pp. 076–083, Dec. 2019, doi: 10.33476/JKY.V27I2.1119.

- [7] H. Febriawati, J. Andri, R. Ali Dori, and U. Muhammadiyah Bengkulu, “Perawatan Infeksi Saluran Pernafasan Akut (ISPA) pada Balita,” *Jurnal Kesmas Asclepius*, vol. 1, no. 1, pp. 25–34, Apr. 2019, doi: 10.31539/JKA.V1I1.526.
- [8] M. T. Hidayatuloh and T. N. Suharsono, “Sistem Pakar Diagnosis Penyakit Infeksi Saluran Pernapasan Akut (ISPA) Menggunakan Metode Dempster Shafer,” *Digital Transformation Technology*, vol. 3, no. 2, pp. 489–498, Oct. 2023, doi: 10.47709/digitech.v3i2.2894.
- [9] J. Simanjuntak, E. Santoso, and M. Marji, “Klasifikasi Penyakit Infeksi Saluran Pernapasan Akut (ISPA) dengan menerapkan Metode Fuzzy K-Nearest Neighbor,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 11, pp. 5023–5029, Oct. 2021, Accessed: Aug. 13, 2024. [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/10163>
- [10] S. S. Indhira and B. Hendrik, “Penerapan Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Ispa Pada Puskesmas Sabak Auh,” *Journal of Information System and Education Development*, vol. 1, no. 2, pp. 6–9, Sep. 2023, Accessed: Aug. 13, 2024. [Online]. Available: <https://journal.mwsfoundation.or.id/index.php/jised/article/view/8>
- [11] Nofitriyani and S. Sunardi, “Perancangan Aplikasi Data Mining Menggunakan Algoritma C4.5 Untuk Mengenali Gejala Penyakit Ispa Pada Balita,” *Jurnal Pertama Indonesia*. Accessed: Aug. 13, 2024. [Online]. Available: <https://jurnal.permataindonesia.ac.id/index.php/JPI/article/view/251>
- [12] M. Napiah, R. A. Purnama, M. Raharjo, and W. Bismi, “Komparasi Algoritma Untuk Klasifikasi Penyakit ISPA (Infeksi Saluran Pernapasan Akut),” Dec. 2022. [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/infortech>
- [13] N. L. W. S. R. Ginantra, I. G. A. D. Indradewi, and E. Hartono, “Machine learning approach for Acute Respiratory Infections (ISPA) prediction: Case study Indonesia,” in *Journal of Physics: Conference Series*, Institute of Physics Publishing, Feb. 2020. doi: 10.1088/1742-6596/1469/1/012044.
- [14] E. Ismanto and M. Novalia, “Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas,” *Techno.Com*, vol. 20, no. 3, pp. 400–410, Aug. 2021, doi: 10.33633/tc.v20i3.4576.
- [15] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, “Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja,” *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 578, Dec. 2021, doi: 10.32493/informatika.v5i4.7575.
- [16] “1. Supervised learning — scikit-learn 1.5.1 documentation.” Accessed: Aug. 17, 2024. [Online]. Available: https://scikit-learn.org/stable/supervised_learning.html
- [17] M. Wati, N. Puspitasari, and A. Sinaga, “Model Algoritma K-Nearest Neighbor Dalam Menganalisis Kinerja Akademik Mahasiswa Teknik Informatika Universitas Mulawarman,” *Jurnal Rekayasa Teknologi Informasi (JURTI)*, vol. 6, no. 2, pp. 124–133, Oct. 2022, Accessed: Aug. 17, 2024. [Online]. Available: <https://e-journals.unmul.ac.id/index.php/INF/article/view/1637>
- [18] A. Damuri, U. Riyanto, H. Rusdianto, and M. Aminudin, “Implementasi Data Mining dengan Algoritma Naïve Bayes Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako,”

- JURIKOM (Jurnal Riset Komputer)*, vol. 8, no. 6, pp. 219–225, Dec. 2021, doi: 10.30865/jurikom.v8i6.3655.
- [19] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *Jurnal Tekno Kompak*, vol. 15, no. 1, pp. 131–145, Feb. 2021, Accessed: Aug. 17, 2024. [Online]. Available: <https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/744>
- [20] E. Saraswati, Y. Umaidah, and A. Voutama, “Penerapan Algoritma Artificial Neural Network untuk Klasifikasi Opini Publik Terhadap Covid-19,” 2021.
- [21] C. Chazar and B. E. Widhiaputra, “Machine Learning Diagnosis Kanker Payudara Menggunakan Algoritma Support Vector Machine,” *Jurnal Informatika dan Sistem Informasi*, vol. 12, no. 1, May 2020.
- [22] F. Yulian Pamuji, V. Puspaning Ramadhan, and R. Artikel, “Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy,” *Jurnal Teknologi dan Manajemen Informatika*, vol. 7, no. 1, pp. 46–50, Jul. 2021, doi: 10.26905/JTMI.V7I1.5982.
- [23] P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, “Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data,” *Ecol Modell*, vol. 406, pp. 109–120, Aug. 2019, doi: 10.1016/j.ecolmodel.2019.06.002.
- [24] “GridSearchCV — scikit-learn 1.5.1 documentation.” Accessed: Aug. 17, 2024. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html
- [25] S. Zhao *et al.*, “Machine Learning Prediction for 50 Anti-Cancer Food Molecules from 968 Anti-Cancer Drugs,” *Int J Intell Sci*, vol. 10, no. 1, pp. 1–8, Jan. 2020, doi: 10.4236/IJIS.2020.101001.
- [26] R. Atallah and A. Al-Mousa, “Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method,” *2019 2nd International Conference on New Trends in Computing Sciences, ICTCS 2019 - Proceedings*, Oct. 2019, doi: 10.1109/ICTCS.2019.8923053.
- [27] W. Musu, A. Ibrahim, and H. Heriadi, “Pengaruh Komposisi Data Training dan Testing terhadap Akurasi Algoritma C4.5,” *SISITI : Seminar Ilmiah Sistem Informasi dan Teknologi Informasi*, vol. 10, no. 1, pp. 186–195, Mar. 2021, Accessed: Aug. 14, 2024. [Online]. Available: <https://www.ejurnal.diponegara.ac.id/index.php/sisiti/article/view/802>
- [28] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 2, no. 5, pp. 178–185, Apr. 2022, doi: 10.30865/KLIK.V2I5.362.
- [29] A. Z. Farmadiansyah, A. F. Hidayatullah, and F. Rahma, “Deteksi Email Spam dan Non Spam Bahasa Indonesia Menggunakan Metode Naïve Bayes,” *AUTOMATA*, vol. 2, no. 2, Aug. 2021, Accessed: Aug. 14, 2024. [Online]. Available: <https://journal.uii.ac.id/AUTOMATA/article/view/19514>