

Terakreditasi SINTA Peringkat 4

Surat Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti No. 28/E/KPT/2019
masa berlaku mulai Vol.3 No. 1 tahun 2018 s.d Vol. 7 No. 1 tahun 2022

Terbit online pada laman web jurnal:
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 7 No. 1 (2022) 35 - 42

JOINTECS

(Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

Deteksi Infeksi *Mycoplasma Pneumoniae Pneumonia* Menggunakan Komparasi Algoritma Klasifikasi *Machine Learning*

Ilham Akhyar Firdaus

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sunan Ampel
ilhamakhyar202@gmail.com

Abstract

Mycoplasma Pneumoniae Pneumonia is a pathogen that attacks the respiratory tract causing an infection. This infection generally occurs in school children and adolescents. Most of these infections are known as mild infections and go away on their own. However, there are cases of extrapulmonary manifestations including neurologic, dermatological, hematological and cardiac syndromes that can result in hospitalization and death. This can be minimized if early detection is carried out on people who are susceptible to the infection. One way is to apply machine learning. So that this can be achieved in this study, several machine learning algorithms will be used, namely Decision Tree, Logistic Regression, Gradient Boosting Decision Tree and Support Vector Machine. Each model will be modified on its hyperparameters using the Grid Search, Random Search and Hyperband methods. The final result shows that the hyperparameter modification method with Hyperband has a slightly better classification performance when compared to Grid Search and Random Search with f-score and accuracy values, namely 0.887 and 0.894 for Decision Tree, 0.942 and 0.947 for Logistic Regression, 0.910 and 0.915 for the Gradient Boosting Decision Tree and 0.591 and 0.715 for Support Vector Machine.

Keywords: machine learning; classification; respiratory tract infection; disease detection; hyperparameter tuning.

Abstrak

Mycoplasma Pneumoniae Pneumonia merupakan salah satu patogen yang menyerang saluran pernapasan sehingga menyebabkan suatu infeksi, Infeksi ini umumnya terjadi pada anak sekolah dan remaja. Sebagian besar infeksi tersebut dikenal sebagai infeksi ringan dan dapat sembuh sendiri. Meskipun begitu, terdapat beberapa kasus yang menyebabkan ekstrapulmoner manifestasi termasuk neurologis, dermatologis, hematologi dan sindrom jantung yang dapat mengakibatkan rawat inap dan kematian. Hal tersebut dapat diminimalisir apabila dilakukan deteksi awal pada orang yang rentan terkena infeksi tersebut. Salah satu caranya adalah dengan menerapkan pembelajaran mesin. Agar hal tersebut dapat dicapai pada penelitian ini akan digunakan beberapa algoritma pembelajaran mesin yaitu *Decision Tree*, *Logistic Regression*, *Gradient Boosting Decision Tree* dan *Support Vector Machine*. Tiap modelnya akan dilakukan modifikasi pada hyper parameternya dengan menggunakan metode Grid Search, Random Search dan Hyperband. Hasil akhir menunjukkan bahwa metode modifikasi hyper parameter dengan Hyperband memiliki performa klasifikasi yang sedikit lebih baik jika dibandingkan dengan *Grid Search* dan *Random Search* dengan nilai f-score dan akurasi masing-masing yaitu 0.887 dan 0.894 untuk *Decision Tree*, 0.942 dan 0.947 untuk *Logistic Regression*, 0.910 dan 0.915 untuk *Gradient Boosting Decision Tree* serta 0.591 dan 0.715 untuk *Support Vector Machine*.

Kata kunci: pembelajaran mesin; klasifikasi; infeksi saluran pernapasan; deteksi penyakit; modifikasi hyperparameter.



1. Pendahuluan

Mycoplasma pneumoniae pneumonia (MPP) merupakan salah satu patogen yang menyerang saluran pernapasan sehingga menyebabkan suatu infeksi. Infeksi ini umumnya terjadi pada anak-anak kecil dan remaja berusia antara 5 sampai 20 tahun [1]. *Mycoplasma pneumoniae pneumonia* juga bertanggung jawab atas sekitar 10-40% kasus *community-acquired pneumonia* (CAP) yang diketahui muncul dengan gejala ringan, namun 3-4% dari mereka dilaporkan berkembang menjadi *pneumonia fulminans* dengan hipoksia dan komplikasi *ekstrapulmoner* yang juga dapat terjadi pada sekitar 25% individu yang terinfeksi MPP [2]. Meskipun *mycoplasma pneumoniae pneumonia* umumnya dianggap sebagai infeksi yang dapat sembuh sendiri, tetapi terkadang dapat menyebabkan berbagai komplikasi paru dan ekstra paru seperti *bronkiolitis obliterans*, *pneumonia nekrotikans*, *ensefalitis*, *arthritis*, *perikarditis*, anemia hemolitik, serta dapat berkembang menjadi *pneumonia berat* yang juga dapat mengancam jiwa [3].

Diagnosis merupakan cara yang dapat dilakukan untuk mendeteksi infeksi tersebut agar dapat ditangani secara cepat dan tepat. Adapun beberapa teknik diagnosis yang dapat dilakukan seperti menggunakan *reverse transcription-polymerase chain reaction* untuk mendeteksi genom virus *pneumonia* [4] ataupun menggunakan pendekatan berbasis data seperti *machine learning* untuk memprediksi apakah seorang pasien terkena *pneumonia* atau tidak dengan menggunakan bantuan alat medis serta merujuk pada data yang lampau [5]. Berdasarkan hal tersebut, maka secara umum terdapat dua teknik yang dapat dilakukan untuk melakukan diagnosis pada infeksi *mycoplasma pneumoniae pneumonia* yaitu menggunakan teknik medis atau pendekatan berbasis data.

Teknik pendekatan berbasis data sendiri juga terbukti dapat dilakukan untuk mendukung pengambilan keputusan dalam dunia kesehatan termasuk mendeteksi sebuah infeksi atau penyakit, hal seperti itu dapat diterapkan dengan menggunakan *machine learning* [6]. Di dalam area kesehatan, *machine learning* digunakan untuk menganalisis parameter klinis seperti hasil tes medis tertentu atau hasil diagnosis medis yang penting dan menggabungkannya untuk memprediksi sebuah penyakit serta dapat menentukan situasi kesehatan pasien secara akurat [7]. *Machine learning* yang dapat melakukan prediksi adalah yang membutuhkan label pada datanya dan umumnya disebut sebagai model *supervised machine learning*. Model *supervised machine learning* yang terbukti dapat melakukan prediksi pada penyakit adalah model klasifikasi seperti *Logistic Regression*, *Decision Tree*, *Gradient Boosting Decision Tree* dan *Support Vector Machine* [8]. Model *Gradient Boosting Decision Tree* dan *Support Vector Machine* memiliki keunggulan yaitu pada *Gradient Boosting Decision Tree* memiliki nilai akurasi tertinggi

dengan waktu komputasi tercepat saat melakukan klasifikasi, sedangkan pada *Support Vector Machine* memiliki waktu komputasi yang efisien secara menyeluruh (tidak hanya tahap klasifikasi) [9]. Selanjutnya pada model *Logistic Regression* dan *Decision Tree* juga memiliki keunggulan masing-masing. *Logistic Regression* dapat memberikan nilai akurasi yang tinggi dengan menghasilkan hasil *training* dan *validation* yang seimbang, sedangkan *Decision Tree* memiliki akurasi yang rendah dibandingkan dengan *Logistic Regression* akan tetapi dapat diaplikasikan pada jenis klasifikasi yang lebih luas dan bervariasi [10].

Berdasarkan pendahuluan tersebut, maka dapat diketahui bahwa infeksi *mycoplasma pneumoniae pneumonia* dapat diminimalisir apabila dilakukan deteksi awal pada orang yang rentan terkena infeksi tersebut. Salah satu caranya adalah dengan menerapkan *machine learning*. Agar hal tersebut dapat dicapai maka pada penelitian ini akan digunakan beberapa algoritma *machine learning* untuk melakukan klasifikasi pada pasien yang rentan terhadap penyakit tersebut dengan melakukan perbandingan performa hasil klasifikasi pada tiap algoritma yang digunakan. Untuk mengetahui perbedaan penelitian ini dengan penelitian sebelumnya maka akan dibandingkan hasil performa klasifikasi dengan menggunakan model *machine learning* yang sama akan tetapi pada penelitian ini akan dilakukan *tuning* pada *hyperparameter* modelnya

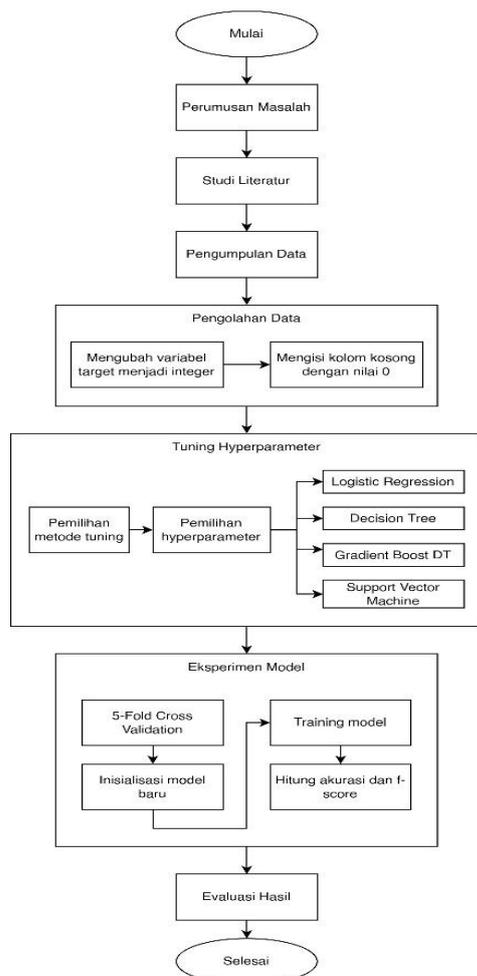
Metode *tuning hyperparameter* merupakan cara untuk menemukan nilai *hyperparameter* yang optimal pada tiap modelnya dengan melihat performa yang dihasilkan oleh model menggunakan matriks evaluasi [11]. Sedangkan *hyperparameter* merupakan istilah dari parameter yang digunakan untuk melakukan konfigurasi pada model sebelum masuk pada proses *training* [12]. Pemilihan *hyperparameter* yang tepat berpengaruh pada performa model. Seperti model *Deep Neural Network* yang memiliki *hyperparameter learning rate* yang dapat mempengaruhi kecepatan gerakan ke arah gradien pada model, sehingga apabila gerakan ke arah gradien terlalu cepat maka proses pelatihan model akan tidak stabil yang mengakibatkan performa model menurun [13]. Dari hal tersebut dapat diketahui bahwa perlu dilakukannya *tuning* pada *hyperparameter* untuk mendapatkan nilai *hyperparameter* yang optimal sehingga performa model dapat mengalami peningkatan.

Pada penelitian ini metode *tuning hyperparameter* yang akan digunakan adalah *Grid Search*, *Random Search*, dan *Hyperband*. Menurut penelitian Xiashuang Wang [14] model yang diberikan metode *Grid Search* memiliki performa tinggi setara dengan *Gradient Boost Machine*. Selanjutnya pada penelitian Tamilvanan [15] *Random Search* mampu mengungguli *Genetic Search* dalam pencarian nilai *hyperparameter* yang optimal. Terakhir pada penelitian Abdul Rehman Khan [16]

Hyperband digunakan untuk melakukan optimasi pada *Gated Recurrent Unit (GRU)* pada *autonomus control* menghasilkan performa yang terbaik dibandingkan dengan GRU saja dan *Long Short Term Memory*. Sehingga pada penelitian ini hal tersebut menjadi landasan untuk melakukan perbandingan performa model dengan menggunakan ketiga metode tuning *hyperparameter* tersebut untuk mencari model dan metode tuning mana yang dapat menghasilkan performa terbaik.

2. Metode Penelitian

Penelitian ini akan menggunakan data pasien yang terkena infeksi *mycoplasma pneumoniae pneumonia* yang selanjutnya akan dilakukan klasifikasi pada data *test*-nya untuk melihat bagaimana hasil performa klasifikasi pada tiap modelnya. Jenis penelitian yang diterapkan yaitu sesuai dengan tujuan penelitian, maka dapat disimpulkan bahwa penelitian ini termasuk dalam penelitian kuantitatif. Hal tersebut dikarenakan berdasarkan jenis data penelitian yang diolah yaitu berupa angka serta analisisnya menggunakan statistik. Metode penelitian yang akan digunakan sesuai dengan Gambar 1 berikut.



Gambar 1. Metode Penelitian

2.1. Studi Literatur

Pada tahapan studi literatur dilakukan pembelajaran dengan mempelajari penelitian-penelitian terdahulu terkait dengan klasifikasi menggunakan *machine learning* terlebih pada studi kasus diagnosis infeksi saluran pernapasan *mycoplasma pneumoniae pneumonia*. Mulai dari pemilihan metode, cara penggunaan metode sampai dengan tahapan evaluasi. Dengan melakukan studi literatur terlebih dahulu diharapkan penelitian dapat dilakukan dengan benar sehingga menghasilkan hasil yang baik.

2.2. Data Penelitian

Data yang digunakan dalam penelitian ini merupakan data pasien yang terkena infeksi *mycoplasma pneumoniae pneumonia*. Terdiri dari 951 baris serta 47 atribut dan 1 label. Semua jenis datanya berupa data numerik. Label yang digunakan menentukan apakah pasien tersebut masuk ke kategori tidak terinfeksi, terinfeksi namun ringan atau terinfeksi namun berat. Tiap label akan dilakukan konversi ke bentuk numerik dengan keterangan seperti pada Tabel 1.

Tabel 1. Konversi Label

Label	Konversi Numerik
Tidak terinfeksi	0
Terinfeksi namun ringan	1
Terinfeksi namun berat	2

Pada label tidak terinfeksi akan dikonversi menjadi angka 0, sedangkan pada label terinfeksi namun ringan akan dikonversi menjadi nomor 1 dan pada label terinfeksi namun berat akan dikonversi menjadi nomor 2. Konversi tersebut dilakukan untuk menyamakan jenis data pada fitur yang lain. Selain itu agar model *machine learning* dapat bekerja lebih baik dan mempercepat dalam proses *training* [17].

2.3. Pengolahan Data

Data yang telah didapatkan perlu diolah terlebih dahulu. Pengolahan data yang dilakukan dalam penelitian ini adalah pertama memastikan bahwa kolom yang menjadi label untuk dilakukan klasifikasi, tipe datanya adalah numerik. Selanjutnya mengisi kolom yang tidak bernilai dengan nilai nol, hal tersebut dilakukan untuk membantu dalam model klasifikasi bekerja dengan lebih baik sehingga hasilnya juga akan ikut baik.

2.4. Modifikasi *Hyperparameter*

Modifikasi *hyperparameter* perlu dilakukan pemilihan metode modifikasinya terlebih dahulu. Metode yang digunakan dalam penelitian ini yaitu dengan *Grid Search*, *Random Search* dan *Hyperband*. Pada metode *tuning* dengan *Grid Search* akan mencari secara lengkap melalui subset yang telah ditentukan secara manual ruang *hyperparameter* dari algoritma pembelajaran dan mengevaluasi setiap kombinasi [18], sehingga *Grid Search* merupakan salah satu metode

tuning yang umum digunakan. Pada metode *Random Search* akan memilih nilai untuk setiap *hyperparameter* secara independen menggunakan distribusi probabilitas dan mengambil nilai *sample* secara *random* [19], dari hal tersebut apabila dimensi fitur berukuran besar maka *Random Search* akan memiliki biaya komputasi yang lebih kecil jika dibandingkan dengan *Grid Search* dan memungkinkan untuk mempunyai nilai *hyperparameter* yang lebih baik [20]. Terakhir pada *Hyperband* merupakan metode *hyper parameter tuning* yang berbasis *bandit-based strategies* dan juga merupakan metode penyempurna dari metode sebelumnya yaitu *Successive Halving* [21].

Berdasarkan penelitian terdahulu yaitu Valarmathi [22], Gunawan [23] dan Eka Patriya [24] ketiga metode tersebut terbukti mampu meningkatkan performa klasifikasi pada model *machine learning*. Terdapat empat model yang selanjutnya akan dilakukan pemilihan *hyper parameter* untuk masing-masing modelnya yaitu *Logistic Regression*, *Decision Tree*, *Gradient Boosting Decision Tree* dan *Support Vector Machine*. Adapun *hyper parameter* dan nilainya yang ingin dioptimalkan pada tiap modelnya yaitu seperti pada Tabel 2.

Tabel 2. Hyperparameter Model

Model	Hyperparameter	Nilai
Logistic Regression	C, Penalty, Solver	(100, 10, 1.0, 0.1, 0.01), ('l1', 'l2', 'elasticnet', 'none'), ('newton-cg', 'lbfgs', 'liblinear', 'sag', 'saga')
Decision Tree	Criterion, Max depth	('gini', 'entropy'), (10, 15, 20, 25, 30)
Gradient Boosting Decision Tree	Learning rate, Estimator	(0.001, 0.01, 0.1, 0.3), (10, 100, 1000)
Support Vector Machine	C	(100, 10, 1.0, 0.1, 0.001)

Pada model pertama yaitu *Logistic Regression* akan dilakukan modifikasi pada nilai *hyperparameter solver*, *penalty* dan C. Nilai *solver* merupakan nilai untuk algoritma yang akan digunakan pada masalah optimasi, sedangkan nilai *penalty* merupakan nilai untuk *penalty* pada logistiknya dan C merupakan nilai regulasi. Selanjutnya pada model kedua yaitu *Decision Tree* akan dilakukan modifikasi pada nilai *hyperparameter criterion* dan *max depth*. Nilai *criterion* merupakan nilai yang mengatur bagaimana nilai *impurity* pada *split* diukur dan pada nilai *max depth* merupakan nilai untuk menentukan kedalaman dari *tree*. Pada model ketiga yaitu *Gradient Boosting Decision Tree* dilakukan pencarian nilai *hyperparameter learning rate* dan *estimator*. *Learning rate* mempengaruhi seberapa cepat model dapat belajar, semakin pelan proses belajarnya maka model akan semakin akurat. Sedangkan pada

nilai estimator merupakan *hyperparameter* yang mempengaruhi dalam banyaknya pohon pada *gradient boost*. Terakhir *Support Vector Machine* adalah nilai C yang merupakan regulasi parameter.

2.5. Eksperimen Model

Pada tahapan eksperimen, model akan dilakukan beberapa proses yaitu pertama proses *training* menggunakan *Grid Search*, *Random Search* dan *Hyperband* dengan melakukan *K-fold cross validation*. Nilai K pada *cross validation* adalah lima, sehingga akan dilakukan lima kali *fold* dengan porsi data *training* dan *testing* masing-masing adalah 80 dan 20. Keluaran dari hasil *training* tersebut adalah nilai *hyperparameter* yang terbaik. Selanjutnya pada proses kedua yaitu melakukan inisialisasi model yang baru dengan memasukkan nilai *hyper parameter* terbaik yang sudah ditemukan menggunakan metode *Grid Search*, *Random Search* dan *Hyperband*. Sehingga selanjutnya pada proses ketiga adalah melakukan *training* model baru. Proses ini bermaksud untuk melatih model dengan data yang sudah disediakan sehingga model dapat melakukan klasifikasi.

Setelah dilakukan *training* pada model baru, selanjutnya menghitung nilai akurasi dan *f-score*. Hal tersebut dilakukan untuk mengetahui bagaimana hasil dari klasifikasi yang dilakukan oleh model tersebut. Adapun rumus dari akurasi yaitu sebagai berikut.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

Rumus 1 digunakan untuk menghitung nilai akurasi sebuah model. TP adalah label dari data yang diprediksi benar dan aslinya benar (*True Positive*). FP merupakan label dari data yang diprediksi benar tetapi aslinya salah (*False Positive*). Variabel TN adalah label dari data yang diprediksi salah dan aslinya salah (*True Negative*). FN yaitu label dari data yang seharusnya diprediksi salah tetapi aslinya benar (*False Positive*). Selanjutnya adalah rumus *f-score* yaitu sebagai berikut.

$$Fscore = 2 * \frac{(Precision*Recall)}{(Precision+Recall)} \quad (2)$$

Rumus 2 digunakan untuk menghitung nilai *f-score* dari sebuah model. Nilai *f-score* merupakan nilai penyeimbang antara *precision* dan *recall* [25]. Dikarenakan hal tersebut *f-score* dapat menjadi matriks penilaian yang lebih baik jika dibandingkan dengan hanya *precision* atau *recall* saja.

2.6. Evaluasi Hasil

Evaluasi hasil merupakan tahapan terakhir dalam penelitian yaitu menarik hasil dan pembahasan dari penelitian menjadi sebuah kesimpulan dan saran. Kesimpulan akan menjawab tujuan dari penelitian dan menjawab permasalahan yang dibahas. Sedangkan pada saran akan diberikan untuk penelitian selanjutnya dengan melihat bagaimana hasil dari penelitian ini dan cara meningkatkan hasilnya dikemudian hari.

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data pasien yang terkena infeksi *mycoplasma pneumoniae pneumoniae*. Terdiri dari 951 baris serta 47 atribut dan satu label. Label yang akan dilakukan klasifikasi terdiri dari tiga label yaitu *no*, *mild* dan *severe* yang masing-masing di representasikan berupa data numerik yaitu 0, 1 dan 2. Keterangannya adalah label *no* menyatakan bahwa pasien tidak terkena infeksi tersebut, sedangkan label *mild* menyatakan bahwa pasien tersebut terkena infeksi akan tetapi ringan dan untuk label *severe* yang merupakan regulasi parameter.

Penelitian ini menggunakan metode *Grid Search*, *Random Search* dan *Hyperband* untuk mencari nilai *hyperparameter* yang optimal pada tiap modelnya. Sedangkan untuk evaluasi pada performa dari model menggunakan nilai *f-score* dan akurasi. Dikarenakan nilai *f-score* dapat menjadi nilai penyeimbang antara *precision* dan *recall* serta nilai akurasi dapat menjadi nilai pembanding dengan *f-score*. Nilai tersebut juga akan dibandingkan dengan hasil penelitian sebelumnya untuk mengetahui seberapa baik hasil penelitian ini jika dibandingkan dengan penelitian sebelumnya. Berikut pada Tabel 3 merupakan hasil serta nilai modifikasi *hyperparameter* beserta *hyperparameter* yang digunakan pada tiap model menggunakan metode *Grid Search*.

Tabel 3. Hasil *Grid Search*

Model	Hyperparameter	Akurasi	F-score
Decision Tree	Criterion: entropy Max depth: 30	0.884	0.865
Logistic Regression	C: 10 Penalty: l2 Solver: newton-cg	0.915	0.903
Gradient Boosting Decision Tree	Learning rate: 0.01 Estimator: 1000	0.926	0.921
Support Vector Machine	C: 0.01 Kernel: linear	0.705	0.581

Tabel 4. Hasil *Random Search*

Model	Hyperparameter	Akurasi	F-score
Decision Tree	Criterion: entropy Max depth: 20	0.831	0.807
Logistic Regression	C: 100 Penalty: l1 Solver: liblinear	0.926	0.920
Gradient Boosting Decision Tree	Learning rate: 0.3 Estimator: 10	0.915	0.910
Support Vector Machine	C: 0.001 Kernel: linear	0.684	0.560

Pada hasil *tuning hyperparameter* menggunakan *Grid Search*, model *Gradient Boosting Decision Tree* memiliki nilai akurasi dan *f-score* yang paling tinggi jika dibandingkan dengan model lainnya. Sedangkan pada model *Support Vector Machine* memiliki nilai

akurasi dan *f-score* terendah jika dibandingkan dengan model lainnya. Sehingga, dengan menggunakan *Grid Search* hampir keseluruhan model selain *Support Vector Machine* memiliki nilai akurasi dan *f-score* yang tinggi. Selanjutnya adalah hasil dari *tuning hyperparameter* dengan menggunakan metode *Random Search* sebagai berikut pada Tabel 4.

Pada hasil *tuning hyperparameter* menggunakan *Random Search* model *logistic Regression* memiliki nilai akurasi dan *f-score* tertinggi jika dibandingkan dengan model lainnya. Sedangkan pada model *Support Vector Machine* tetap menjadi model yang memiliki nilai akurasi dan *f-score* terendah. Jika dibandingkan dengan *Grid Search*, *Random Search* memiliki nilai akurasi dan *f-score* terendah seperti pada ketiga modelnya yaitu *Decision Tree*, *Gradient Boosting Decision Tree* dan *Support Vector Machine*. Selanjutnya pada Tabel 5 adalah hasil *tuning* menggunakan *Hyperband* sebagai berikut.

Tabel 5. Hasil *Hyperband*

Model	Hyperparameter	Akurasi	F-score
Decision Tree	Criterion: entropy Max depth: 15	0.894	0.887
Logistic Regression	C: 1.0 Solver: newton-cg	0.947	0.942
Gradient Boosting Decision Tree	Learning rate: 0.3 Estimator: 10	0.915	0.910
Support Vector Machine	C: 0.001 Kernel: linear	0.715	0.591

Tabel 6. Hasil Penelitian Sebelumnya

Model	Hyperparameter	Akurasi	F-score
Decision Tree	-	0.873	0.852
Logistic Regression	Penalty: l2	0.694	0.589
Gradient Boosting Decision Tree	Estimator: 300	0.926	0.920
Support Vector Machine	Kernel: linear	0.652	0.478

Pada hasil *tuning hyperparameter* menggunakan *Hyperband* model *logistic Regression* juga memiliki nilai akurasi dan *f-score* tertinggi jika dibandingkan dengan model lainnya. Sedangkan pada model *Support Vector Machine* tetap menjadi model yang memiliki nilai akurasi dan *f-score* terendah. Jika dibandingkan dengan *Grid Search* serta *Random Search*, *Hyperband* unggul dalam pemilihan nilai *hyperparameter* pada model *Logistic Regression*, *Decision Tree* dan *Support Vector Machine*. Sedangkan pada model *Gradient Boosting Decision Tree* tidak lebih baik dari *Grid Search*. Ketiga hasil tersebut selanjutnya akan dibandingkan dengan penelitian sebelumnya untuk mengetahui bagaimana perbedaan hasil performa klasifikasi dengan menggunakan model *machine learning* yang sama akan tetapi pada penelitian ini dilakukan *tuning* pada *hyperparameter* modelnya. Diharapkan dengan melakukan *tuning hyperparameter*

dapat memberikan performa klasifikasi yang lebih baik. Berikut pada Tabel 6 merupakan hasil penelitian sebelumnya.

Pada penelitian yang dilakukan oleh Chenglin Pan, dkk yang berjudul *Artificial Intelligence Enhanced Rapid and Efficient Diagnosis of Mycoplasma Pneumoniae Pneumonia in Children Patients*. Dengan menerapkan model yang sama akan tetapi pada penelitian ini dilakukan tuning pada *hyperparameter* tiap modelnya yang diharapkan performa dari model yang ada sebelumnya dapat mengalami peningkatan. Jika dibandingkan semua model dengan ketiga metode tuning tersebut. Tuning menggunakan *hyperband* memiliki nilai akurasi dan *f-score* yang lebih tinggi pada ketiga modelnya selain model *Gradient Boosting Decision Tree* dengan masing-masing kenaikan model yaitu untuk *Decision Tree* sebesar 0.035, *Logistic Regression* sebesar 0.353 dan *Support Vector Machine* sebesar 0.113.

Pada metode *Grid Search* terdapat kenaikan nilai pada seluruh modelnya baik pada nilai akurasi maupun *f-score* nya. Akan tetapi jika dibandingkan dengan *hyperband* maka hanya model *Gradient Boosting Decision Tree* saja yang memiliki nilai yang lebih tinggi. Saat proses komputasi jika dibandingkan dengan kedua metode lainnya, *Grid Search* juga cenderung lebih lama. Hal tersebut dikarenakan pada *Grid Search* seluruh nilai akan dicoba satu persatu sehingga apabila semakin banyak nilai *Hyperparameter* yang dicari maka biaya komputasinya juga akan semakin banyak.

Selanjutnya pada metode *Random Search* terdapat beberapa model yang justru mengalami penurunan yaitu *Decision Tree* dan *Gradient Boosting Decision Tree* jika dibandingkan dengan *Grid Search* dan *Hyperband*. Sedangkan pada model *Logistic Regression* dan *Support Vector Machine* mengalami peningkatan baik nilai *f-score* maupun akurasi jika dibandingkan dengan *Grid Search*. Meskipun *Random Search* cenderung lebih cepat jika dibandingkan dengan *Grid Search*, akan tetapi hasil pencarian nilai *hyperparameter* pada *Random Search* juga belum tentu optimal bila dibandingkan dengan *Grid Search*.

4. Kesimpulan

Penelitian ini membuktikan bahwa dalam melakukan deteksi infeksi *mycoplasma pneumoniae pneumonia* dapat dilakukan dengan menerapkan model *machine learning*. Dari keempat model yang diusulkan, model *Logistic Regression* memiliki performa klasifikasi yang lebih baik jika dibandingkan dengan ketiga model lainnya dengan melakukan *tuning* pada *hyperparameter*nya. Selanjutnya pada metode *tuning hyperparameter* dengan *Hyperband* merupakan metode *tuning* yang melakukan pemilihan *hyperparameter* terbaik jika dibandingkan dengan kedua metode lainnya yaitu *Grid Search* dan *Random Search* dengan model *machine learning* berturut-turut yaitu *Decision Tree*,

Logistic Regression dan *Support Vector Machine*.

Saran pada penelitian selanjutnya yaitu melakukan analisis lebih dalam pada tiap *hyperparameter* serta melihat metode atau cara lain seperti dengan memilih fitur yang penting dari data menggunakan *feature selection* untuk meningkatkan performa klasifikasi dari modelnya. Adapun dalam penelitian ini tidak memperhatikan apakah kelas dari data seimbang atau tidak dikarenakan untuk menyeimbangkan hasil yang didapat pada penelitian sebelumnya yang mana tidak memperhatikan apakah kelas dari datanya seimbang atau tidak. Sehingga pada penelitian selanjutnya dapat memperhatikan apakah kelas dari data seimbang atau tidak, apabila tidak seimbang maka dapat menggunakan metode *undersampling* atau *oversampling* untuk mengatasinya.

Daftar Pustaka

- [1] A. Dierig, H. H. Hirsch, M. L. Decker, J. A. Bielicki, U. Heininger, and N. Ritz, "Mycoplasma pneumoniae detection in children with respiratory tract infections and influence on management – a retrospective cohort study in Switzerland," *Acta Paediatr. Int. J. Paediatr.*, vol. 109, no. 2, pp. 375–380, 2020.
- [2] S. H. Yoon, I. K. Min, and J. G. Ahn, "Immunochromatography for the diagnosis of Mycoplasma pneumoniae infection: A systematic review and meta-analysis," *PLoS One*, vol. 15, no. 3, pp. 1–17, 2020.
- [3] Y. Zhang, Y. Zhou, S. Li, D. Yang, X. Wu, and Z. Chen, "The clinical characteristics and predictors of refractory Mycoplasma pneumoniae pneumonia in children," *PLoS One*, vol. 11, no. 5, pp. 1–10, 2016.
- [4] M. S. Han *et al.*, "Contribution of Co-detected Respiratory Viruses and Patient Age to the Clinical Manifestations of Mycoplasma Pneumoniae Pneumonia in Children," *Pediatr. Infect. Dis. J.*, vol. 37, no. 6, pp. 531–536, 2018.
- [5] Y. H. Liao, Z. C. Wang, F. G. Zhang, M. F. Abbod, C. H. Shih, and J. S. Shieh, "Machine learning methods applied to predict ventilator-associated pneumonia with pseudomonas aeruginosa infection via sensor array of electronic nose in intensive care unit," *Sensors (Switzerland)*, vol. 19, no. 8, 2019.
- [6] J. Tohka and M. van Gils, "Evaluation of machine learning algorithms for health and wellness applications: A tutorial," *Comput. Biol. Med.*, vol. 132, no. March, p. 104324, 2021.
- [7] T. T. Han *et al.*, "Machine learning based classification model for screening of infected patients using vital signs," *Informatics Med. Unlocked*, vol. 24, p. 100592, 2021.
- [8] Y. W. Lee, J. W. Choi, and E. H. Shin, "Machine learning model for diagnostic method prediction in parasitic disease using clinical information,"

- Expert Syst. Appl.*, vol. 185, no. February, p. 115658, 2021.
- [9] C. Zhang, C. Liu, X. Zhang, and G. Almpanidis, “An up-to-date comparison of state-of-the-art classification algorithms,” *Expert Syst. Appl.*, vol. 82, pp. 128–150, 2017.
- [10] W. Chen *et al.*, “Groundwater spring potential mapping using artificial intelligence approach based on kernel logistic regression, random forest, and alternating decision tree models,” *Appl. Sci.*, vol. 10, no. 2, pp. 1–23, 2020.
- [11] J. Wu, X. Y. Chen, H. Zhang, L. D. Xiong, H. Lei, and S. H. Deng, “Hyperparameter optimization for machine learning models based on Bayesian optimization,” *J. Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019.
- [12] L. Yang and A. Shami, “On hyperparameter optimization of machine learning algorithms: Theory and practice,” *Neurocomputing*, vol. 415, pp. 295–316, 2020.
- [13] Y. J. Yoo, “Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches,” *Knowledge-Based Syst.*, vol. 178, pp. 74–83, 2019.
- [14] X. Wang, G. Gong, and N. Li, “Automated recognition of epileptic EEG states using a combination of symlet wavelet processing, gradient boosting machine, and grid search optimizer,” *Sensors (Switzerland)*, vol. 19, no. 2, 2019.
- [15] T. Tamilvanan and V. Murali Bhaskaran, “A New Feature Selection Techniques Using Genetics Search and Random Search Approaches For Breast Cancer,” *Biosci. Biotechnol. Res. Asia*, vol. 14, no. 1, pp. 409–414, 2017.
- [16] A. Rehman Khan, A. Tamoor Khan, M. Salik, and S. Bakhsh, “An Optimally Configured HP-GRU Model Using Hyperband for the Control of Wall Following Robot,” *Int. J. Robot. Control Syst.*, vol. 1, no. 1, pp. 66–74, 2021.
- [17] A. Ambarwari, Q. J. Adrian, and Y. Herdiyeni, “Analisis Pengaruh Data Scaling Terhadap Performa Algoritme Machine Learning untuk Identifikasi Tanaman,” *Rekayasa Sist. dan Teknol. Inf.*, vol. 4, no. 1, pp. 117–122, 2020.
- [18] R. Ghawi and J. Pfeffer, “Efficient Hyperparameter Tuning with Grid Search for Text Categorization using kNN Approach with BM25 Similarity,” *Open Comput. Sci.*, vol. 9, no. 1, pp. 160–180, 2019.
- [19] A. Panichella, “A Systematic Comparison of search-Based approaches for LDA hyperparameter tuning,” *Inf. Softw. Technol.*, vol. 130, p. 106411, 2021.
- [20] R. Andonie and A. C. Florea, “Weighted random search for CNN hyperparameter optimization,” *Int. J. Comput. Commun. Control*, vol. 15, no. 2, pp. 1–11, 2020.
- [21] L. Li, K. Jamieson, and G. Desalvo, “Hyperband : A Novel Bandit-Based Approach to,” vol. 18, pp. 1–52, 2018.
- [22] R. Valarmathi and T. Sheela, “Heart disease prediction using hyper parameter optimization (HPO) tuning,” *Biomed. Signal Process. Control*, vol. 70, no. July, p. 103033, 2021.
- [23] M. I. Gunawan, D. Sugiarto, and I. Mardianto, “Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode Grid Search pada Algoritma Logistic Regression,” *JEPIN (Jurnal Edukasi dan Penelit. Inform.)*, vol. 6, no. 3, pp. 280–284, 2020.
- [24] E. Eka Patriya, “Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (Ihsg),” *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 24–38, 2020.
- [25] A. Alomari, N. Idris, A. Q. M. Sabri, and I. Alsmadi, “Deep reinforcement and transfer learning for abstractive text summarization: A review,” *Comput. Speech Lang.*, vol. 71, no. August 2021, p. 101276, 2022.

Halaman ini sengaja dikosongkan