

Analisis Prediksi *Length of Stay* Pasien Infeksi Paru Menggunakan Algoritma Klasifikasi

<http://dx.doi.org/10.28932/jutisi.v12i1.11464>

Riwayat Artikel

Received: 14 Maret 2025 | Final Revision: 8 April 2026 | Accepted: 8 April 2026

Creative Commons License 4.0 (CC BY – NC)



Glory Emilisa Rupilu^{✉#1}, Swat Lie Liliawati^{#2}, Mewati Ayub^{#3}

[#] Program Magister Ilmu Komputer, Universitas Kristen Maranatha
Jl. Prof. drg. Surya Sumantri, M.P.H. No. 65 Bandung-40164, Indonesia

¹2379006@maranatha.ac.id

²swat.ll@maranatha.ac.id

³mewati.ayub@maranatha.ac.id

✉ Corresponding author: 2379006@maranatha.ac.id

Abstrak — Penelitian ini menginvestigasi hubungan antara faktor usia, jenis kelamin, dan factor lain dalam atribut dengan *Length of Stay* (LOS) pada pasien dengan penyakit paru. Tujuan utama penelitian adalah untuk membantu memprediksi *length of stay* pasien baru yang datang dengan diagnosa yang sama dan untuk membantu mengurangi biaya perawatan yang berkaitan dengan durasi tinggal di rumah sakit. Teori yang digunakan dalam penelitian ini adalah faktor usia, jenis kelamin, diagnosa, nilai leukosit dan hasil rontgen dada dapat mempengaruhi durasi tinggal mereka di rumah sakit. Data untuk penelitian ini diperoleh dari rekam medis salah satu rumah sakit di Jawa Barat selama periode penelitian selama kurang lebih 3 bulan. Metode dan teknik yang digunakan adalah *Artificial Neural Network-Multilayer Perceptron*, naïve bayes, J48 dan *Random Tree* untuk menganalisis dan memodelkan hubungan antara variabel input (usia, jenis kelamin, diagnosa sekunder dan lain-lain) dan variabel output. Hasil dari penelitian ini diharapkan dapat memberikan pemahaman yang lebih baik tentang faktor-faktor yang mempengaruhi *length of stay* pada pasien dengan penyakit paru, serta memberikan kontribusi dalam pengembangan metode prediksi yang dapat membantu manajemen pasien dan pengambilan keputusan klinis yang lebih baik di rumah sakit.

Kata kunci— J48; LOS; Multilayer perceptron; Naïve bayes; Random Tree

Length of Stay Prediction Analysis for Pulmonary Infection Patients using Classification Algorithms

Abstract — This study investigates the relationship between age, gender, and other factors in attributes with *Length of Stay* (LOS) in patients with pulmonary disease. The main objective of the study was to help predict the length of stay new patients presenting with the same diagnosis and to help reduce the cost of care related to the duration of hospital stay. The theory used in this study is that the factors of age, gender, diagnosis, leukocyte values and chest X-ray results can affect the duration of their stay in the hospital. Data for this study was obtained from the medical records of one of the hospitals in West Java during the study period for approximately 3 months. The methods and techniques used are *Artificial Neural Network-Multilayer Perceptron*, naïve bayes, J48 and *Random Tree* to analyze and model the relationship between input variables (age, gender, secondary diagnoses and others) and output variable. The results of this study are expected to provide a better understanding of the factors that influence length of stay in patients with pulmonary diseases, as well as contribute to the development of prediction methods that can help better patient management and clinical decision-making in hospitals.

Keywords— *J48; LOS; Multilayer perceptron; Naïve bayes; Random Tree*

I. PENDAHULUAN

Data WHO tahun 2019, pneumonia menyebabkan 14% dari seluruh kematian anak di bawah 5 tahun dengan total kematian 740.180 jiwa. Data Riskesdas Indonesia tahun 2018, menyatakan penderita pneumonia meningkat seiring dengan bertambahnya usia. Pada kelompok usia 55-64 tahun mencapai 2,5%, pada kelompok usia 65-74 tahun sebesar 3,0% dan pada kelompok usia 75 tahun keatas mencapai 2,9% tertuang pada Artikel Kemenkes tahun 2022. Efisiensi proses perawatan di rumah sakit menjadi hal penting dalam meningkatkan kualitas pelayanan kesehatan. Dimensi mutu pelayanan kesehatan di Indonesia disepakati mengacu pada 7 dimensi yang digunakan oleh WHO dan lembaga internasional yang mana, salah satunya adalah efisien dalam mengoptimalkan pemanfaatan sumber daya yang tersedia dan mencegah pemborosan termasuk alat kesehatan, obat, energi dan ide yang tercantum pada Permenkes no 30, 2022. Salah satu indikator penilaian mutu pelayanan rawat inap di RS adalah rerata jumlah hari pasien dirawat di Rumah Sakit. Sejumlah penelitian sebelumnya yang menyelidiki faktor-faktor yang mempengaruhi Length of stay (LOS) pasien pneumonia mengidentifikasi variabel-variabel yang berhubungan dengan pasien seperti usia lanjut dan komorbiditas tertentu, selain tingkat keparahan penyakit, sebagai prediktor lamanya perawatan di rumah sakit yang berkepanjangan [1]. Lama perawatan pasien di Rumah Sakit dipengaruhi oleh banyak faktor diantaranya keadaan klinis saat pertama kali masuk RS, perawatan selama sakit baik pengobatan dan tindakan medis serta manajemen pasien. LOS menunjukkan berapa hari lama perawatan seorang pasien dalam satu jangka waktu [2]. Sebuah model yang didasarkan pada penyematan dan Feed-Forward Neural Network (FFNN) oleh Vincent et al. adalah FFNN lebih baik dalam memprediksi LOS dibandingkan dengan random forest atau regresi logistik [3]. Pneumonia adalah peradangan akut jaringan paru yang disebabkan oleh mikroorganisme seperti virus, bakteri dan jamur yang dapat menimbulkan gejala ringan hingga berat bahkan kematian pada semua kelompok usia di dunia [4]. Bronkopneumonia adalah peradangan pada saluran pernapasan yang terjadi di bronkus hingga alveolus paru, masalah ini sering terjadi pada anak kecil [5].

Hasil yang ingin dicapai adalah melihat hubungan yang signifikan antara usia, jenis kelamin, diagnosa sekunder, antibiotik, nilai leukosit, kesimpulan rontgen dada dan LOS pada penyakit paru. Metode pembelajaran yang dilakukan menggunakan *neural network* analisis dengan *multilayer perceptron* yang kemudian dibandingkan dengan *naive bayes* serta *J48* dan *random tree*. Data yang diambil merupakan data nyata dari sebuah Rumah Sakit swasta dalam waktu 5 bulan pada pasien rawat inap yang pulang dengan penyakit *pneumonia* dan *bronkopneumonia*. Aplikasi yang digunakan dalam menganalisa data adalah WEKA. WEKA adalah singkatan dari *Waikato Environment for Knowledge Analysis* merupakan aplikasi gratis untuk melakukan berbagai tugas pembelajaran mesin dan penggalian data sebagai bagian dari aktivitas harian pengolah data.

II. METODE PENELITIAN

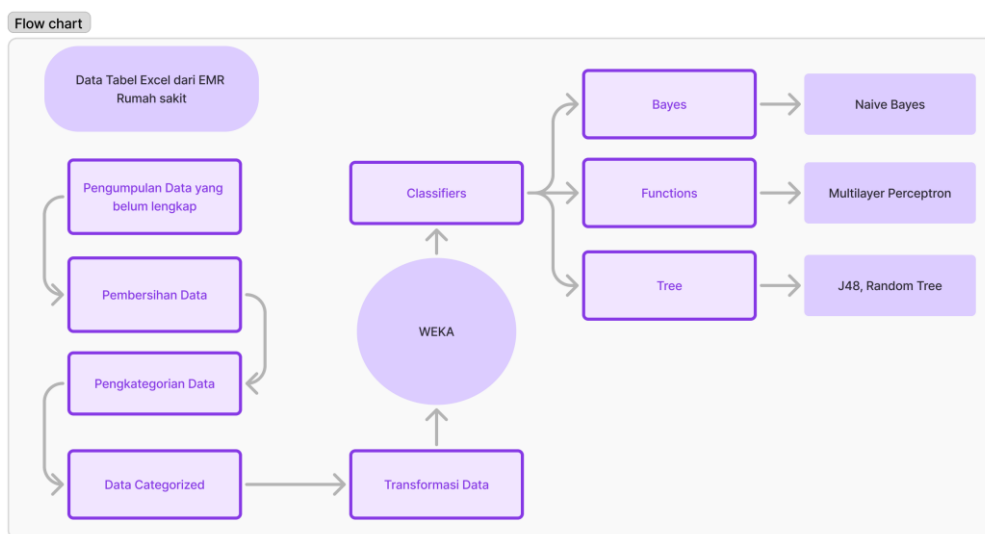
Neural network adalah sebuah algoritma untuk pembelajaran mesin dengan dua dasar kategori yaitu regresi dan klasifikasi. *Neural network* pertama kali muncul pada awal tahun 1940-an dimana jaringan saraf ini terdiri dari tiga lapisan unit saraf yaitu lapisan unit *input* dengan lapisan unit *hidden* kemudian lapisan unit *output*. ANN untuk memprediksi penyakit virus, kanker, kulit, dan COVID-19 juga mencapai akurasi lebih dari 97% dan presisi 98% pada penelitian Azeem et al. [6]. Beberapa tipe *Neural Network* antara lain adalah *feed-forward neural network*, *cascade-neural network*, *fit net shallow neural network*, *shallow neural network*, *convolutional neural network*, dan *recursive neural network* [7]. *Artificial Neural Network* (ANN) atau Jaringan Saraf Tiruan merupakan sistem permodelan yang dibuat dengan pola hubungan rumit antara *input* dan *output* [7]. Jaringan saraf tiruan telah dikembangkan untuk menyelidiki hubungan nonlinear yang kompleks antara variabel prediktor dan parameter yang diprediksi menghasilkan *Multi-layered Perception* (MLP) sebagai salah satu arsitektur ANN yang sukses [8]. *Perceptron multilayer* (MLP) adalah jenis jaringan saraf umpan maju yang terdiri dari neuron yang terhubung penuh dengan jenis fungsi aktivasi nonlinier namun perceptron ini banyak digunakan untuk membedakan data yang tidak dapat dipisahkan secara linier. Fleksibilitas MLP dalam arsitektur dan kemampuannya untuk memperkirakan fungsi apa pun dalam kondisi tertentu menjadikannya penyusun mendasar dalam pembelajaran mendalam dan penelitian jaringan saraf.

Naïve Bayes adalah algoritma yang terkenal untuk klasifikasi probabilistik. Ini adalah algoritma yang sederhana tetapi efektif yang dapat digunakan untuk berbagai aplikasi [9] yang ada di dunia nyata, seperti memberikan rekomendasi produk, melakukan diagnosis medis, dan mengontrol kendaraan otonom. Ada variasi *Naïve Bayes* yang tersedia untuk memenuhi data umum karena kegagalan data nyata yang memenuhi asumsi *Naïve Bayes* [10]. Penelitian Blanquero et al. menemukan dibandingkan dengan pendekatan seleksi fitur yang memiliki referensi yang baik, *Naïve Bayes* jarang yang diusulkan mendapatkan hasil yang kompetitif dalam hal akurasi, sparsitas, dan waktu berjalan untuk set data yang seimbang [11]. Secara umum penelitian yang dilakukan Ayyouzadeh et al. pada faktor yang mempengaruhi lamanya perawatan adalah faktor jumlah layanan paramedis, frekuensi konsultasi, bangsal klinis, spesialisasi dan tingkat pendidikan dokter, serta penyebab rawat inap.

Hasil penelitian mereka terlihat bahwa sensitivitas terbaik (80,36%) masing-masing dimiliki oleh Regresi Logistik dan *Naive Bayes* dan nilai AUC terbaik (0,896) dimiliki oleh klasifikasi *Random Forest* dan *Generalized Linear* [12].

Lingkup pembelajaran WEKA digunakan dengan menguji pendekatan pohon keputusan J48 dan *random tree*. J48 adalah implementasi *open-source* Java dari algoritma C4.5 (pengklasifikasi statistik). Pohon keputusan J48 dibangun secara iteratif, satu simpul pada satu waktu. Setiap simpul dalam pohon tersebut mewakili sebuah klasifikasi dan cabang-cabang yang menghubungkan simpul tersebut dengan simpul akar adalah kondisi-kondisi yang menghasilkan klasifikasi. Pohon keputusan transparan yang berbeda dapat dibandingkan dan dievaluasi sesuai dengan akurasi prediktif masing-masing untuk memastikan model memprediksi dengan benar kelas dari data yang baru atau data yang tidak terlihat [13]. *Random tree* termasuk klasifikasi algoritma yang lebih cepat diproses [14].

Pemrosesan data yang dilakukan tergambar pada Gambar 1 yang meringkaskan pengolahan pemodelan data pasien pulang rawat inap dengan diagnosa penyakit paru dalam kurun waktu 5 bulan yang diambil dari *electronic medical record* salah satu rumah sakit di Jawa Barat. Data yang diperoleh dilengkapi dan dibersihkan kemudian dikategorikan dengan fitur jenis kelamin, kategori usia, kategori hari rawat, hari rawat, umur tahun, diagnosa utama lain, diagnosa sekunder, leukosit, jenis antibiotik dan kategori rontgen dada. Target kelas yang dituju adalah kategori hari rawat yang dibagi menjadi 3 yaitu cepat, sedang dan lama.



Gambar 1. Pengolahan pemodelan data

III. HASIL DAN PEMBAHASAN

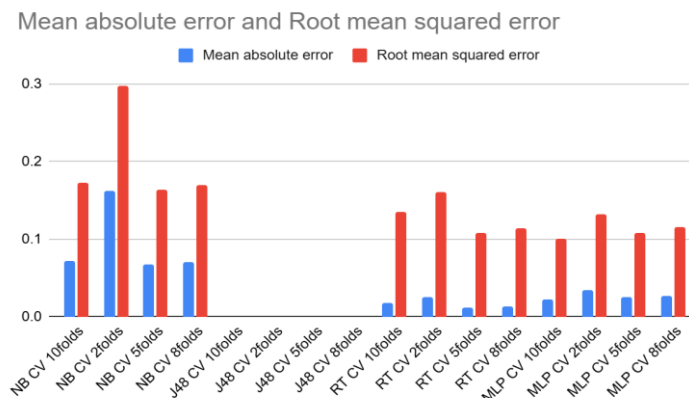
Data yang telah dikumpulkan dan dibersihkan kemudian disimpan dalam bentuk CSV yang diolah menggunakan aplikasi WEKA dengan algoritma *machine learning* model MLP, *naive bayes*, J48 dan *random tree*. *Neural network* yang dipakai pada pembelajaran ini adalah *multilayer perceptron* yang mampu memperkirakan fungsi dalam kondisi tertentu pada *deep learning* dan penelitian jaringan saraf. Pada Tabel 1 terdapat hasil data yang diproses dengan klasifikasi MLP dengan pilihan *test cross validation folds-10* dan *percentage split 66%* menunjukkan hasil model bahwa *cross validation folds-10* dengan *default tuning parameter* memiliki tingkat akurasi model lebih tinggi dibanding 3 tuning folds lainnya. Proses pembagian data disebut *k-fold cross validation*, yang membantu mencegah *overfitting* dengan melakukan resampling dataset berdasarkan parameter yang disebut *k* [15], [16] namun setelah dilakukan pemrosesan data hasil akurasi yang didapatkan diatas 90%. Tipe error 1 dan 2 pada klasifikasi yang digunakan cenderung memberikan tingkat akurasi yang sempurna, namun *naive bayes* dan *random tree* memiliki tingkat kesalahan yang muncul. *Naive Bayes* (NB) dapat gagal dalam memprediksi contoh minoritas [17] karena sensitivitasnya terhadap distribusi kelas yang cenderung pada kelas dominan demikian pula *random tree* memiliki variasi yang acak menyebabkan kurangnya sensitivitas terhadap kelas minoritas. Gambar 2 dan Gambar 3 meringkaskan perbandingan MAE dan RMSE *cross validation k-folds* dan *percentage split* yang menyimpulkan *naive bayes* memiliki tingkat akurasi yang kurang karena nilai MAE dan RMSE yang lebih tinggi dari model yang lain.

TABEL 1
HASIL MODEL DENGAN TEST OPSI CROSS VALIDATION K-FOLDS

| | NB CV 10folds | NB CV 2folds | NB CV 5folds | NB CV 8folds | J48 CV 10folds | J48 CV 2folds | J48 CV 5folds | J48 CV 8folds |
|----------------------------------|---------------|--------------|--------------|--------------|----------------|---------------|---------------|---------------|
| Correctly Classified Instances | 94.9268 % | 80.2927 % | 95.2195 % | 94.9268 % | 100 % | 100 % | 100 % | 100 % |
| Incorrectly Classified Instances | 5.0732 % | 19.7073 % | 4.7805 % | 5.0732 % | 0 % | 0 % | 0 % | 0 % |
| Kappa statistic | 0.9088 | 0.6418 | 0.9141 | 0.9088 | 1 | 1 | 1 | 1 |
| Mean absolute error | 0.0713 | 0.1629 | 0.0679 | 0.0706 | 0 | 0 | 0 | 0 |
| Root mean squared error | 0.172 | 0.298 | 0.1643 | 0.1692 | 0 | 0 | 0 | 0 |
| Relative absolute error | 19.4615 % | 44.468 % | 18.5347 % | 19.2687 % | 0 % | 0 % | 0 % | 0 % |
| Root relative squared error | 40.2013 % | 69.6559 % | 38.4001 % | 39.5632 % | 0 % | 0 % | 0 % | 0 % |

TABEL 2
HASIL MODEL DENGAN TEST OPSI CROSS VALIDATION K-FOLDS

| | RT CV 10folds | RT CV 2folds | RT CV 5folds | RT CV 8folds | MLP CV 10folds | MLP CV 2folds | MLP CV 5folds | MLP CV 8folds |
|----------------------------------|---------------|--------------|--------------|--------------|----------------|---------------|---------------|---------------|
| Correctly Classified Instances | 971707% | 961951% | 982439% | 980488% | 954146% | 940488% | 949268% | 940488% |
| Incorrectly Classified Instances | 28293% | 38049% | 17561% | 19512% | 45854% | 59512% | 50732% | 59512% |
| Kappa statistic | 0.9485 | 0.9304 | 0.9681 | 0.9644 | 0.9141 | 0.8872 | 0.9044 | 0.8872 |
| Mean absolute error | 0.0186 | 0.026 | 0.0117 | 0.013 | 0.0217 | 0.0341 | 0.0247 | 0.027 |
| Root mean squared error | 0.1357 | 0.1601 | 0.1082 | 0.1141 | 0.101 | 0.1322 | 0.1087 | 0.1153 |
| Relative absolute error | 50747% | 70919% | 31968% | 35522% | 5939% | 93132% | 67316% | 73752% |
| Root relative squared error | 317156% | 374269% | 252951% | 266634% | 236216% | 309108% | 254209% | 26966% |



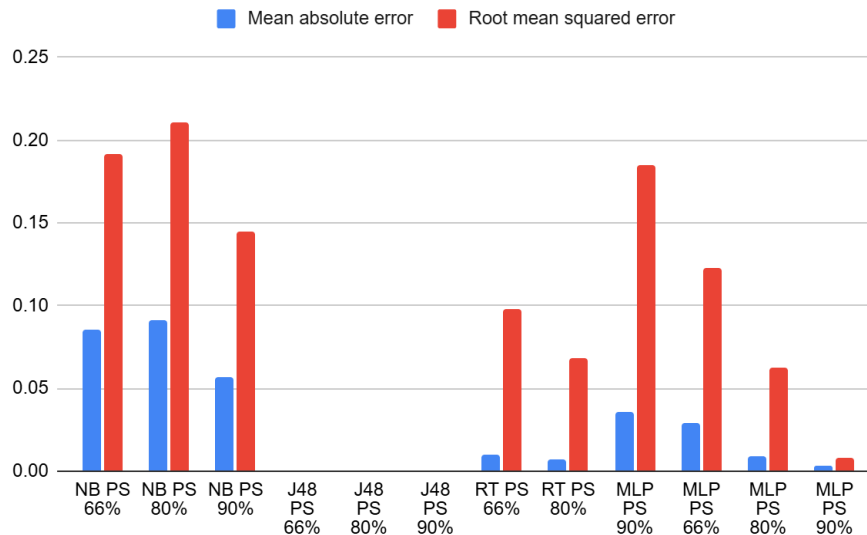
Gambar 2. Perbandingan MAE dan RMSE cross validation k-folds

TABEL 3
HASIL MODEL DENGAN OPSI *PERCENTAGE SPLIT*

| | NB PS 66% | NB PS 80% | NB PS 90% | J48 PS 66% | J48 PS 80% | J48 PS 90% |
|---|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|
| Correctly Classified Instances | 93.9655 % | 92.6829 % | 96.0784 % | 100 % | 100 % | 100 % |
| Incorrectly Classified Instances | 6.0345 % | 7.3171 % | 3.9216 % | 0 % | 0 % | 0 % |
| Kappa statistic | 0.8922 | 0.8661 | 0.9267 | 1 | 1 | 1 |
| Mean absolute error | 0.0859 | 0.0915 | 0.0569 | 0 | 0 | 0 |
| Root mean squared error | 0.1914 | 0.2108 | 0.1447 | 0 | 0 | 0 |
| Relative absolute error | 23.4739 % | 25.1516 % | 15.829 % | 0 % | 0 % | 0 % |
| Root relative squared error | 44.8052 % | 49.7529 % | 34.5556 % | 0 % | 0 % | 0 % |

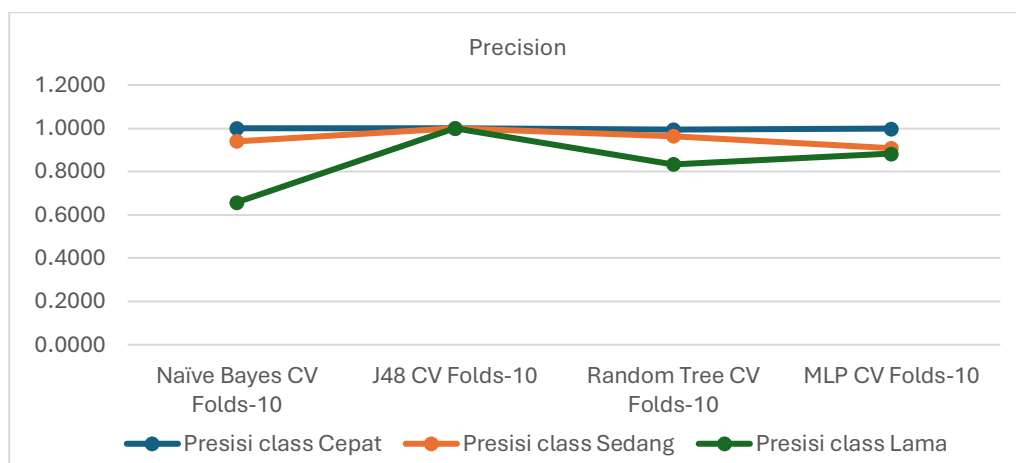
TABEL 4.
HASIL MODEL DENGAN OPSI *PERCENTAGE SPLIT*

| | RT PS 66% | RT PS 80% | RT PS 90% | MLP PS 66% | MLP PS 80% | MLP PS 90% |
|---|----------------------|----------------------|----------------------|-----------------------|-----------------------|-----------------------|
| Correctly Classified Instances | 985632% | 995122% | 95098% | 942529% | 995122% | 100% |
| Incorrectly Classified Instances | 14368% | 0.4878 % | 4902% | 57471% | 0.4878 % | 0% |
| Kappa statistic | 0.9735 | 0.9909 | 0.9058 | 0.8908 | 0.9909 | 1 |
| Mean absolute error | 0.0096 | 0.0069 | 0.0359 | 0.0289 | 0.0091 | 0.0032 |
| Root mean squared error | 0.0979 | 0.068 | 0.1852 | 0.1225 | 0.0624 | 0.0076 |
| Relative absolute error | 26165% | 19005% | 100039% | 78886% | 25098% | 0.8848 % |
| Root relative squared error | 22909% | 160514% | 442481% | 286766% | 147218% | 18177% |



Gambar 3. Perbandingan MAE dan RMSE percentage split

Penggunaan WEKA dalam mengolah data yang diperoleh menggunakan metode klasifikasi *naïve bayes*, J48 dan pohon acak pada pengaturan awal dengan metode tes *cross validation 10-folds* dan *percentage split 66%* menghasilkan data seperti pada Tabel 1 dan 2. Pada Tabel 1,2,3, dan 4 model-model ini memiliki tingkat akurasi paling tinggi 100% pada J48 dan paling rendah pada *naïve bayes* berarti model cukup baik dalam mengklasifikasikan data dengan nilai paling tinggi adalah 6.0345 % dari total instance yang salah klasifikasi. Nilai Kappa terendah pada *naïve bayes* dengan nilai 0.8908, menunjukkan tingkat kesepakatan yang sangat tinggi antara prediksi model dengan data sebenarnya yang menandakan bahwa model jauh lebih baik daripada tebakan acak. *Mean Absolute Error* (MAE) mengkarakterisasi perubahan di antara nilai asli dan nilai yang dapat diprediksi dan ditambah sebagai rata-rata perubahan total set data [18] demikian dapat digunakan untuk mengukur seberapa dekat dengan perkiraan seperti pada pengembangan model berbasis data untuk prakiraan kecepatan angin di Australia [19]. Mengevaluasi kesalahan model dengan *Mean Absolute Error* (MAE) paling sedikit pada J48 yang berarti kesalahan rata-rata kecil. *Root Mean Square Error* (RMSE) adalah ukuran kinerja yang didefinisikan sebagai akar kuadrat dari ekspektasi selisih kuadrat antara nilai yang diestimasi dan nilai actual dimana ukuran ini digunakan untuk menilai keakuratan estimasi parameter [20]. *Root Mean Squared Error* (RMSE) paling kecil pada J48 menyatakan kesalahan masih dalam batas yang dapat diterima. *Relative Absolute Error* (RAE) paling tinggi pada *naïve bayes* model *test percentage split 66%* yaitu 23.47% dan *Root Relative Squared Error* (RRSE) 44.80%, menunjukkan bahwa model cukup baik dibandingkan dengan baseline. Pada Gambar 4 terlihat presisi model class lama rawat inap cenderung lebih kurang pada *naïve bayes*, kemudian model class sedang paling rendah pada MLP dan class cepat cenderung menghasilkan model yang lebih presisi.



Gambar 4. Perbandingan Presisi pada *Naïve Bayes*, J48, *Random Tree* dan MLP dengan *test option CV Folds-10*

Berdasarkan data analisis model *Multilayer Perceptron* (MLP) *cross validation k-folds*, faktor usia, jenis kelamin, nilai leukosit, dan diagnosa sekunder mempengaruhi *Length of Stay* (LOS) pada pasien dengan penyakit infeksi paru. Kategori usia memiliki bobot yang berbeda untuk setiap kelompok terhadap LOS dimana Bayi dan Balita memiliki bobot negatif yang lebih besar pada node tertentu, menunjukkan bahwa mereka cenderung memiliki LOS yang lebih pendek. Lansia memiliki bobot positif yang lebih tinggi pada beberapa node, menunjukkan bahwa mereka cenderung memiliki LOS yang lebih lama. Faktor usia juga berkontribusi secara langsung melalui atribut UmurTahun, di mana bobot positif menunjukkan bahwa semakin tua usia pasien, semakin besar kemungkinan LOS meningkat. Hal ini dapat ditarik kesimpulan bahwa Usia merupakan faktor signifikan dalam menentukan LOS. Pasien lansia cenderung membutuhkan waktu rawat inap yang lebih lama dibandingkan pasien berusia muda. Bobot untuk jenis kelamin perempuan (JK=P) bervariasi di setiap node, tetapi secara umum kontribusinya kecil dibandingkan dengan faktor lain seperti usia atau jumlah hari rawat. Dalam beberapa node, jenis kelamin perempuan memiliki bobot negatif kecil, menunjukkan bahwa perempuan mungkin sedikit lebih cepat pulih dibandingkan laki-laki. Dapat ditarik kesimpulan bahwa jenis kelamin memiliki pengaruh yang relatif kecil terhadap LOS dibandingkan faktor lain seperti usia atau kondisi medis. Nilai leukosit memiliki bobot positif pada beberapa node, menunjukkan bahwa peningkatan kadar leukosit berhubungan dengan LOS yang lebih lama. Hal ini sesuai dengan fakta medis bahwa kadar leukosit yang tinggi biasanya menunjukkan infeksi atau inflamasi yang lebih parah, sehingga membutuhkan waktu pemulihan yang lebih lama. Dapat ditarik kesimpulan bahwa nilai leukosit adalah indikator penting untuk menentukan tingkat keparahan infeksi paru dan memprediksi LOS. Diagnosa sekunder memiliki bobot negatif kecil di sebagian besar node. Ini menunjukkan bahwa keberadaan diagnosa sekunder sedikit meningkatkan kemungkinan LOS menjadi lebih lama. Pasien dengan diagnosa sekunder mungkin membutuhkan perawatan tambahan atau lebih kompleks, sehingga memperpanjang waktu rawat inap. Dapat disimpulkan bahwa diagnosa sekunder berkontribusi terhadap peningkatan LOS meskipun pengaruhnya tidak sebesar faktor lain seperti usia atau leukosit. Atribut "HariRawat" memiliki bobot paling signifikan di hampir semua node baik positif maupun negatif tergantung pada konteksnya. Bobot positif besar menunjukkan bahwa jumlah hari rawat sangat menentukan klasifikasi pasien ke kategori "Cepat", "Sedang", atau "Lama". Kesimpulan yang dapat ditarik adalah hari rawat adalah variabel utama dalam menentukan LOS karena secara langsung mencerminkan durasi perawatan pasien.

| Attribute | Class | | |
|--------------|-----------------|------------------|----------------|
| | Cepat (0.51) | Sedang (0.43) | Lama (0.06) |
| ===== | | | |
| JK | | | |
| L | 288.0 | 209.0 | 33.0 |
| P | 241.0 | 232.0 | 28.0 |
| [total] | 529.0 | 441.0 | 61.0 |
| KategoriUsia | | | |
| Bayi | 1.0 | 2.0 | 1.0 |
| Balita | 207.0 | 145.0 | 11.0 |
| Anak | 92.0 | 54.0 | 2.0 |
| Remaja | 72.0 | 54.0 | 5.0 |
| Dewasa | 122.0 | 106.0 | 21.0 |
| Lansia | 39.0 | 84.0 | 25.0 |
| [total] | 533.0 | 445.0 | 65.0 |

Gambar 5. Model *naive bayes* dengan *cross validation 10-folds*

Pada Gambar 5 distribusi klasifikasi untuk model *naive bayes* pada jenis kelamin Laki-laki (L) terdapat 288 pasien pada kategori rawat Cepat, 209 pasien pada kategori rawat sedang, 33 pasien pada kategori rawat lama, kemudian Perempuan (P) terdapat 241 pasien dengan kategori rawat cepat, 232 pasien dengan kategori rawat sedang, dan 28 pasien dengan kategori rawat lama. Laki-laki cenderung lebih banyak diklasifikasikan sebagai "Cepat" dibandingkan perempuan. Namun, perempuan memiliki jumlah yang lebih tinggi di kategori "Sedang". Ini menunjukkan bahwa jenis kelamin dapat mempengaruhi LOS, di mana laki-laki lebih cepat pulih daripada perempuan dalam konteks ini. Dari data pemodelan *naive bayes* dapat dilihat bahwa kelompok usia balita dan dewasa memiliki jumlah yang signifikan dalam kategori "Cepat". Sebaliknya, lansia memiliki proporsi lebih tinggi dalam kategori "Sedang" dan "Lama". Ini menunjukkan bahwa usia

berpengaruh besar terhadap LOS; pasien yang lebih tua cenderung memerlukan waktu lebih lama untuk pulih. Peningkatan kadar leukosit biasanya menunjukkan adanya infeksi atau inflamasi. Rata-rata yang lebih tinggi pada kategori "Lama" menunjukkan bahwa pasien dengan infeksi yang lebih parah (ditunjukkan oleh kadar leukosit yang tinggi) cenderung mengalami LOS yang lebih panjang. Keberadaan diagnosa sekunder tampak berkontribusi pada peningkatan LOS. Pasien dengan kondisi kesehatan tambahan mungkin memerlukan perawatan yang lebih intensif dan waktu pemulihan yang lebih lama. Analisis pada model J48 menunjukkan bahwa faktor utama yang mempengaruhi LOS adalah jumlah hari rawat. Analisis pengaruh faktor terhadap lama rawat inap dengan model *Random Tree*, terlihat yang berpengaruh adalah rata-rata leukosit yang lebih tinggi pada kategori inap "Lama" menunjukkan bahwa pasien dengan infeksi yang lebih parah memerlukan waktu pemulihan yang lebih lama. Hal ini mengingatkan untuk merekomendasikan identifikasi pasien dengan kadar leukosit tinggi sejak awal perawatan untuk memberikan intervensi yang lebih agresif dan memantau perkembangan mereka secara lebih intensif.

IV. SIMPULAN

Model *Multilayer Perceptron* (MLP) menunjukkan performa yang baik dalam memprediksi LOS berdasarkan berbagai atribut pasien. Dengan menggunakan *cross-validation 10-fold*, MLP dapat menangkap hubungan kompleks antar variabel yang mempengaruhi LOS. Faktor-Faktor yang Mempengaruhi LOS antara lain usia, jenis kelamin, nilai leukosit, diagnosa sekunder dan hari rawat. Kategori usia memiliki bobot yang berbeda, di mana Bayi dan Balita cenderung memiliki LOS yang lebih pendek, sementara Lansia memiliki LOS yang lebih lama. Ini menunjukkan bahwa usia adalah faktor signifikan dalam menentukan LOS. Meskipun bobot untuk jenis kelamin perempuan bervariasi, kontribusinya terhadap LOS relatif kecil dibandingkan dengan faktor lain seperti usia dan hari rawat. Kadar leukosit yang tinggi berhubungan dengan LOS yang lebih lama, menunjukkan bahwa nilai ini adalah indikator penting untuk memprediksi tingkat keparahan infeksi paru. Keberadaan diagnosa sekunder sedikit meningkatkan kemungkinan LOS menjadi lebih lama, meskipun pengaruhnya tidak sebesar faktor usia atau leukosit. Atribut "HariRawat" memiliki bobot paling signifikan dalam menentukan LOS. Jumlah hari rawat secara langsung mencerminkan durasi perawatan pasien dan sangat menentukan klasifikasi pasien ke dalam kategori "Cepat", "Sedang", atau "Lama". Secara keseluruhan, model MLP terbukti efektif dalam memprediksi LOS dengan mempertimbangkan beberapa faktor klinis. Dengan pemahaman mendalam tentang bagaimana faktor-faktor ini mempengaruhi durasi rawat inap, rumah sakit dapat meningkatkan manajemen pasien dan kualitas pelayanan kesehatan secara keseluruhan.

Hasil analisis ini masih memiliki keterbatasan, terutama karena sifat data yang retrospektif, jumlah sampel terbatas, dan belum dilakukannya validasi eksternal sehingga generalisasi model perlu diuji lebih lanjut. Analisis prediksi selanjutnya diharapkan melakukan validasi eksternal pada dataset rumah sakit lain atau dataset publik seperti MIMIC [21], mengembangkan model *ensemble* dan multimodal dengan menggabungkan data citra rontgen dan data klinis untuk meningkatkan akurasi, serta mengintegrasikan model ke dalam sistem informasi manajemen rumah sakit (SIMRS) dalam bentuk *dashboard* yang mudah diinterpretasikan. Selain itu, diperlukan penerapan teknik mitigasi *overfitting* seperti *pruning*, regularisasi, dan validasi prospektif agar model lebih baik sebelum diterapkan dalam praktik klinis.

DAFTAR PUSTAKA

- [1] G. Lüthi-Corridori *et al.*, 'Predictors of Length of Stay, Rehospitalization and Mortality in Community-Acquired Pneumonia Patients: A Retrospective Cohort Study', *J. Clin. Med.*, vol. 12, no. 17, 2023.
- [2] H. Purawijaya, Y. P. Satar, N. Andarusito, E. Hadimuljono, and R. Ruahedi, 'Analisis Faktor-Faktor Yang Mempengaruhi Length Of Stay (LOS) di Instalasi Gawat Darurat Rumah Sakit Hermina Ciputat', *J. Manaj. dan Adm. Rumah Sakit Indones.*, vol. 7, no. 4, pp. 356–368, 2023.
- [3] V. Lequertier, T. Wang, J. Fondrevelle, V. Augusto, S. Polazzi, and A. Duclos, 'Length of Stay Prediction With Standardized Hospital Data From Acute and Emergency Care Using a Deep Neural Network', *Med. Care*, vol. 62, no. 4, pp. 225–234, 2024.
- [4] A. Torres *et al.*, 'Pneumonia', *Nat. Rev. Dis. Prim.*, vol. 7, no. 1, 2021.
- [5] Y. Ling, D. Yang, and S. Yang, 'Clinical characteristics, early blood biochemical indicators, and prognostic status of children with bronchopneumonia', *Medicine (Baltimore)*, vol. 102, no. 47, p. E36162, 2023.
- [6] M. Azeem *et al.*, 'Neural Networks for the Detection of COVID-19 and Other Diseases: Prospects and Challenges', *Bioeng. (Basel, Switzerland)*, vol. 10, no. 7, 2023.
- [7] R. Qamar and B. A. Zardari, 'Artificial Neural Networks: An Overview', *Mesopotamian J. Comput. Sci.*, vol. 2023, pp. 124–133, 2023.
- [8] S. R. Shams, A. Jahani, S. Kalantary, M. Moenaddini, and N. Khorasani, 'The evaluation on artificial neural networks (ANN) and multiple linear regressions (MLR) models for predicting SO2 concentration', *Urban Clim.*, vol. 37, p. 100837, 2021.
- [9] S. A. Shubho, M. R. H. Razib, N. K. Rudro, A. K. Saha, M. S. U. Khan, and S. Ahmed, 'Performance analysis of NB Tree, REP tree and random tree classifiers for credit card fraud data', *2019 22nd Int. Conf. Comput. Inf. Technol. ICCIT 2019*, 2019.
- [10] I. Wickramasinghe and H. Kalutarage, 'Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation', *Soft Comput.*, vol. 25, no. 3, pp. 2277–2293, 2021.
- [11] R. Blanco, E. Carrizosa, P. Ramirez-Cobo, and M. R. Sillero-Denamiel, 'Variable selection for Naïve Bayes classification', *Comput. Oper. Res.*, vol. 135, p. 105456, 2021.
- [12] S. M. Ayyoubzadeh *et al.*, 'A study of factors related to patients' length of stay using data mining techniques in a general hospital in southern Iran', *Heal. Inf. Sci. Syst.*, vol. 8, no. 1, 2020.
- [13] M. Singh, P. K. Gupta, V. Tyagi, J. Flusser, T. Ören, and G. Valentino, Eds., 'Advances in Computing and Data Sciences', vol. 1244, 2020.
- [14] D. R. Hermawan, M. Fahrio Ghaniah Fatimah, L. Kurniawati, and A. Helen, 'Comparative Study of J48 Decision Tree Classification Algorithm,

- Random Tree, and Random Forest on In-Vehicle Coupon Recommendation Data', *2021 Int. Conf. Artif. Intell. Big Data Anal. ICAIBDA 2021*, pp. 76–81, 2021.
- [15] R. M. Azad San Ahmed and R. M. Azad San, 'Hard Voting Approach using SVM, Naïve Bays and Decision Tree for Kurdish Fake News Detection', *Iraqi J. Comput. Sci. Math.*, vol. 4, no. 3, p. 3, 2023.
- [16] P. Charilaou and R. Battat, 'Machine learning models and over-fitting considerations', *World J. Gastroenterol.*, vol. 28, no. 5, pp. 605–607, 2022.
- [17] T. Kim and J. S. Lee, 'Maximizing AUC to learn weighted naive Bayes for imbalanced data classification', *Expert Syst. Appl.*, vol. 217, p. 119564, 2023.
- [18] A. S. Rajawat, O. Mohammed, R. N. Shaw, and A. Ghosh, 'Renewable energy system for industrial internet of things model using fusion-AI', *Appl. AI IOT Renew. Energy*, pp. 107–128, 2022.
- [19] A. Neupane, N. Raj, R. Deo, and M. Ali, 'Development of data-driven models for wind speed forecasting in Australia', *Predict. Model. Energy Manag. Power Syst. Eng.*, pp. 143–190, 2021.
- [20] S. K. Jain and V. P. Singh, 'Statistical Techniques for Data Analysis', in *Developments in Water Science*, vol. 51, 2003, pp. 207–276.
- [21] T. Zebin, S. Rezvy, and T. J. Chausalet, 'A deep learning approach for length of stay prediction in clinical settings from medical records', *2019 IEEE Conf. Comput. Intell. Bioinforma. Comput. Biol. CIBCB 2019*, 2019.