Walisongo Journal of Information Technology
Vol 8 No 1 (2024): 31-40. DOI: 10.21580/wjit.2026.8.1.22179
ISSN 2715-0143 (online) | 2714-9048 (printed)

Pengembangan Model Prediksi Risiko Tuberkulosis Paru Menggunakan Algoritma CatBoost di Puskesmas Kabangka, Kabupaten Muna

Rina[1], Elvi Saktiawati Salemaku[2], Anya Salsabila[3], Rizal Adi Saputra[4]
[123] Teknik Informatika, Universitas Halu Oleo
Corresponding Author : waoderina75@gmail.com

Abstract: *Pulmonary tuberculosis (TB) remains a major public health concern, particularly in Indonesia, where the burden of the disease remains high. Although the Directly Observed Treatment Shortcourse (DOTS) strategy has been implemented to control TB, the detection of new cases remains a challenge. This study aims to develop a pulmonary TB risk prediction model using the CatBoost algorithm based on patient data from Kabangka Community Health Center, Muna Regency. A total of 2,724 patient records were used in the analysis. The research process involved data preprocessing, dataset splitting, model training, and performance evaluation using Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and Root Mean Squared Error (RMSE). The evaluation results yielded an MAE of 0.00467, an MSE of 0.00385, a MAPE of 36.37%, and an RMSE of 0.062. These results indicate the prediction errors obtained from the implemented model and provide an initial assessment of its performance on the dataset. The model is expected to support preliminary analysis of pulmonary TB risk and provide information that may assist public health planning and interventions.*

Keyword: *CatBoost, pulmonary tuberculosis, risk prediction, machine learning, Puskesmas Kabangka.*

Copyright © 2026 Walisongo Journal of Information Technology

Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

1. Introduction

Tuberkulosis (TB) paru merupakan salah satu penyakit menular yang masih menjadi permasalahan kesehatan masyarakat di Indonesia. Pengendalian TB memerlukan upaya berkelanjutan melalui penemuan kasus, pengobatan yang tepat, serta identifikasi kelompok yang berisiko. Strategi Directly Observed Treatment Shortcourse (DOTS) telah diterapkan untuk mendukung pengendalian TB, tetapi upaya penemuan kasus baru masih menghadapi berbagai tantangan. Oleh karena itu, pemanfaatan data pasien untuk menganalisis pola dan faktor yang berkaitan dengan TB menjadi salah satu pendekatan yang dapat dikaji guna mendukung perencanaan program kesehatan.

Perkembangan machine learning membuka peluang untuk mengolah data kesehatan dan membangun model prediksi berdasarkan pola yang terdapat dalam dataset. Penelitian (Ula et

al., 2023) menerapkan K-Means dan regresi linear untuk menganalisis tingkat risiko TB paru berdasarkan data wilayah. (Kurniawan & Salam, 2024) mengkaji seleksi fitur Information Gain untuk mendukung klasifikasi penyakit tuberkulosis. Sementara itu, penelitian mengenai klasifikasi TB menggunakan algoritma C4.5 dan Extreme Learning Machine menunjukkan adanya pendekatan komputasional lain yang dapat digunakan untuk mengolah data terkait penyakit tersebut. Berbagai penelitian ini memperlihatkan bahwa pemanfaatan algoritma pembelajaran mesin telah menjadi salah satu pendekatan dalam analisis data TB, meskipun setiap penelitian memiliki karakteristik data, tujuan, dan metode evaluasi yang berbeda.

Penerapan machine learning pada data pasien juga telah dikaji untuk mendukung analisis kondisi klinis penderita tuberkulosis. (Nurul Abdillah & Alfita Dewi, 2024), misalnya, mengembangkan aplikasi berbasis web dengan algoritma Naïve Bayes untuk memprediksi status kegawatan pasien tuberkulosis berdasarkan 227 data rekam medis. Penelitian tersebut menggunakan karakteristik pasien dan gejala klinis sebagai atribut prediktor, dengan hasil akurasi sebesar 71,81%. Selain itu, (Nurhaliza et al., 2025) menerapkan Random Forest yang dikombinasikan dengan teknik penyeimbangan data ADASYN dan Tomek Links untuk klasifikasi tuberkulosis. Penelitian tersebut menyoroti pentingnya penanganan ketidakseimbangan kelas dalam meningkatkan kemampuan model mengenali kasus positif. (Ramadhani et al., 2025) juga mengkaji klasifikasi data pasien TB paru menggunakan Probabilistic Neural Network pada data dari Puskesmas Telaga Sari, Kota Balikpapan. Penelitian-penelitian tersebut menunjukkan bahwa pemilihan algoritma, karakteristik data, dan tahapan pengolahan data merupakan aspek penting dalam pengembangan model berbasis machine learning untuk analisis tuberkulosis.

Selain klasifikasi berdasarkan data pasien, pendekatan pembelajaran mesin juga digunakan untuk menganalisis pola risiko pada tingkat wilayah. (Ramadhani et al., 2025) mengintegrasikan machine learning dan analisis spasial untuk memprediksi wilayah rawan tuberkulosis di Provinsi Sumatera Utara. Penelitian tersebut menggunakan variabel yang mencakup kepadatan penduduk, status gizi, ketersediaan fasilitas kesehatan, tingkat kemiskinan, kualitas hunian, dan cakupan imunisasi. Pendekatan ini menunjukkan bahwa analisis berbasis data dapat dimanfaatkan untuk mendukung pemetaan risiko dan perencanaan intervensi kesehatan. Namun, analisis risiko wilayah berbeda dari prediksi berdasarkan karakteristik individu, sehingga metode dan interpretasi hasil perlu disesuaikan dengan tujuan penelitian.

Salah satu algoritma yang dapat digunakan dalam pemodelan prediktif adalah CatBoost (*Categorical Boosting*), yaitu algoritma berbasis gradient boosting yang dirancang untuk menangani fitur kategorikal. CatBoost telah diterapkan pada berbagai permasalahan prediksi di luar bidang tuberkulosis. (Putri & Arianto, 2024), misalnya, membandingkan Random Forest, CatBoost, dan XGBoost untuk memprediksi konsentrasi PM10 di Jakarta Utara. (Syukur et al., 2025) menerapkan CatBoost untuk memprediksi kualitas udara, sedangkan (Eko Fitra Firmandani et al., 2024) mengkaji kombinasi CatBoost, seleksi fitur, dan hyperparameter tuning untuk prediksi nasabah potensial. Penelitian lainnya juga mengkaji penggunaan CatBoost untuk klasifikasi indeks standar pencemaran udara dengan penanganan ketidakseimbangan kelas (Aditya et al., 2025). Penelitian-penelitian tersebut memberikan gambaran tentang penerapan CatBoost dalam pemodelan prediktif, tetapi hasilnya tidak dapat langsung digeneralisasi untuk prediksi risiko TB paru.

Berdasarkan uraian tersebut, penelitian ini bertujuan mengembangkan model prediksi risiko penyakit TB paru menggunakan algoritma CatBoost dengan memanfaatkan data pasien dari Puskesmas Kabangka, Kabupaten Muna. Dataset yang digunakan terdiri atas 2.724 data pasien dengan variabel yang mencakup umur, jenis kelamin, status perkawinan, pendidikan, pekerjaan, akses terhadap air bersih, perilaku merokok, serta kondisi kesehatan, termasuk hipertensi dan gejala batuk berkepanjangan. Tahapan penelitian meliputi preprocessing data,

pembagian dataset, pelatihan model, dan evaluasi hasil prediksi. Penelitian ini diharapkan dapat memberikan gambaran mengenai penerapan CatBoost pada data pasien tingkat puskesmas sebagai pendekatan analitis yang berpotensi mendukung perencanaan program pengendalian TB paru.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif dengan menerapkan algoritma Categorical Boosting (CatBoost) untuk membangun model prediksi berdasarkan data pasien Puskesmas Kabangka, Kabupaten Muna. Tahapan penelitian meliputi pengumpulan data, preprocessing, pembagian dataset, pelatihan model, dan evaluasi kinerja. Penyusunan tahapan tersebut mengacu pada prinsip pengembangan model pembelajaran mesin, yaitu mempersiapkan data, melatih model, dan mengevaluasi hasil prediksi menggunakan metrik yang sesuai. Penelitian terdahulu mengenai klasifikasi penyakit tuberkulosis juga menunjukkan pemanfaatan pembelajaran mesin untuk menganalisis data pasien dan memilih fitur yang relevan (Kurniawan & Salam, 2024)

Gambar 1. Diagram Alir Metodologi



2.1 Dataset

Data penelitian diperoleh dari Puskesmas Kabangka, Kabupaten Muna, dengan jumlah sebanyak 2.724 data pasien. Variabel yang digunakan mencakup umur, jenis kelamin, status perkawinan, pendidikan, pekerjaan, akses terhadap air bersih, perilaku merokok, serta kondisi kesehatan berupa hipertensi dan gejala batuk berkepanjangan. Variabel target yang tercantum dalam dataset adalah kolom “DI DIAGNOSIS TB PARU”.

Data tersebut digunakan sebagai dasar pembentukan model prediksi menggunakan algoritma CatBoost. Pemilihan variabel dan penentuan target perlu mempertimbangkan karakteristik data serta tujuan pemodelan. (Kurniawan & Salam, 2024) , misalnya, mengkaji seleksi fitur *Information Gain* untuk mendukung klasifikasi penyakit tuberkulosis. Penelitian (Ula et al., 2023) juga membahas pemanfaatan metode pembelajaran mesin untuk analisis tingkat risiko TB paru.

2.2 Preprocessing

Tahap *preprocessing* dilakukan untuk mempersiapkan data sebelum digunakan dalam pemodelan. Proses yang tercatat dalam penelitian ini meliputi pembersihan data, penyeimbangan kelas, seleksi fitur, dan *encoding*. Pembersihan data dilakukan sebagai bagian dari persiapan dataset. Penyeimbangan kelas digunakan untuk menangani distribusi kelas yang tidak seimbang, sedangkan seleksi fitur bertujuan menentukan fitur yang digunakan dalam pemodelan. Selanjutnya, *encoding* dilakukan untuk menyesuaikan representasi data kategorikal dengan format yang diperlukan dalam proses pengolahan data.

Penanganan fitur kategorikal merupakan salah satu karakteristik penting algoritma CatBoost. (Prokhorenkova et al., 2018) menjelaskan bahwa CatBoost menggunakan mekanisme pemrosesan fitur kategorikal dan *ordered boosting* untuk mengatasi permasalahan *prediction shift* pada proses pembelajaran. Selain itu, penelitian (Kurniawan & Salam, 2024) memperlihatkan penerapan seleksi fitur dalam pemodelan klasifikasi penyakit tuberkulosis. Dalam penelitian ini, teknik spesifik yang digunakan untuk pembersihan data, penyeimbangan kelas, seleksi fitur, dan *encoding* mengikuti implementasi yang telah dilakukan.

2.3 Pembagian Data

Setelah proses *preprocessing*, dataset dibagi menjadi data pelatihan dan data pengujian. Data pelatihan digunakan untuk membangun model, sedangkan data pengujian digunakan untuk mengukur kinerja model terhadap data yang tidak digunakan secara langsung selama pelatihan. Pembagian ini bertujuan memperoleh evaluasi yang lebih representatif terhadap kemampuan model menghasilkan prediksi pada data pengujian.

Pembagian data merupakan bagian penting dalam pengembangan model pembelajaran mesin karena hasil evaluasi perlu didasarkan pada data yang terpisah dari proses pelatihan. (Ihzaniah et al., 2023), misalnya, membandingkan metode regresi menggunakan metrik kesalahan prediksi untuk mengevaluasi kinerja model. Dalam penelitian ini, pembagian dataset dilakukan sebelum pelatihan dan pengujian model. Rasio pembagian dan teknik pembagian yang digunakan perlu dicantumkan berdasarkan konfigurasi eksperimen yang sebenarnya.

2.4 Pelatihan Model

Pemodelan dilakukan menggunakan algoritma CatBoost, yaitu algoritma *gradient boosting* berbasis pohon keputusan yang dirancang untuk menangani fitur kategorikal. CatBoost membangun model secara bertahap melalui proses *boosting*. (Prokhorenkova et al., 2018) menjelaskan bahwa algoritma ini menggabungkan mekanisme *ordered boosting* dan pemrosesan fitur kategorikal untuk mengurangi permasalahan *prediction shift* yang dapat muncul selama pelatihan model.

Parameter yang disebutkan dalam penelitian ini meliputi *learning rate*, *depth*, dan *iterations*. Parameter *learning rate* mengatur besarnya kontribusi pembaruan model pada setiap tahap pembelajaran, *depth* menentukan kedalaman pohon keputusan, sedangkan *iterations* menentukan jumlah iterasi *boosting*. Pengaturan parameter tersebut menjadi bagian dari proses pengembangan model. (Syandika & Yustanti, 2023) juga menerapkan CatBoost dengan optimisasi hiperparameter untuk memprediksi pembatalan pesanan kamar hotel.

Setelah konfigurasi model ditentukan, pelatihan dilakukan menggunakan data pelatihan. Model yang dihasilkan kemudian digunakan untuk memprediksi data pengujian. Untuk menjaga keterulangan penelitian, nilai parameter yang digunakan dan prosedur pemilihannya perlu dilaporkan sesuai dengan hasil eksperimen yang sebenarnya.

2.5 Evaluasi

Evaluasi dilakukan untuk mengukur perbedaan antara nilai prediksi dan nilai aktual. Penelitian ini menggunakan empat metrik evaluasi, yaitu *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), *Mean Absolute Percentage Error* (MAPE), dan *Root Mean Squared Error* (RMSE). Metrik-metrik tersebut digunakan dalam evaluasi model regresi, sebagaimana dibahas dalam penelitian (Ihzaniah et al., 2023) .

1) Mean Absolute Error (MAE)

MAE merupakan rata-rata nilai absolut selisih antara nilai aktual dan nilai prediksi. Metrik ini menunjukkan besarnya kesalahan prediksi rata-rata.

Rumus:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}|$$

Ketarangan:
yi: Nilai sebenarnya.
y: Nilai prediksi.
N: Jumlah observasi.

2) Mean Squared Error (MSE)

MSE merupakan rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi. Karena kesalahan dikuadratkan, kesalahan yang lebih besar memberikan kontribusi lebih besar terhadap nilai MSE.

Rumus:

$$MSE = \frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y})^2$$

Keterangan:
N: Jumlah sampel dalam dataset
yi: Nilai sebenarnya dari sampel ke-i
y: Nilai prediksi dari model untuk sampel ke-i.

3) Mean Absolute Percentage Error (MAPE)

MAPE mengukur rata-rata kesalahan absolut relatif terhadap nilai aktual dalam bentuk persentase.

Rumus:

$$MAPE = \frac{\sum_{t=1}^{n} \left| \left(\frac{A_t - F_t}{A_t} \right) 100 \right|}{n}$$

Keterangan
At: Aktual permintaan ke t
Ft: Hasil peramalan ke t
N: Besarnya data peramalan

MAPE perlu diinterpretasikan secara hati-hati apabila terdapat nilai aktual nol karena pembagian dengan nol tidak terdefinisi.

4) Root Mean Squared Error (RMSE)
RMSE merupakan akar kuadrat dari MSE. Metrik ini menyatakan besarnya kesalahan prediksi dalam satuan yang sama dengan nilai target.

Rumus:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2}$$

Keterangan:
yi: Nilai sebenarnya.
y: Nilai prediksi.
N: Jumlah observasi.

Pada persamaan tersebut, n merupakan jumlah observasi, yi adalah nilai aktual observasi ke-i, sedangkan y_prediksi merupakan nilai prediksi model untuk observasi ke-i. Hasil perhitungan keempat metrik digunakan untuk menggambarkan kinerja model sebagaimana dilaporkan pada bagian hasil dan pembahasan.

3. Hasil dan Pembahasan

3.1 Pengumpulan data

Penelitian ini menggunakan data pasien Puskesmas Kabangka, Kabupaten Muna, sebanyak 2.724 data untuk membangun model prediksi risiko tuberkulosis (TB) paru menggunakan algoritma CatBoost. Data yang digunakan mencakup sejumlah variabel karakteristik pasien dan faktor yang berkaitan dengan kondisi kesehatan, termasuk perilaku merokok dan gejala batuk berdahak selama dua minggu atau lebih. Dataset tersebut selanjutnya digunakan dalam proses pemodelan untuk menghasilkan nilai prediksi yang kemudian dievaluasi menggunakan metrik kesalahan prediksi.

Tabel 1 menyajikan contoh struktur data yang digunakan dalam penelitian. Untuk menjaga kerahasiaan informasi pasien, nama pada tabel ditampilkan menggunakan kode anonim. Penyajian ini tetap memperlihatkan struktur variabel yang dianalisis tanpa mengungkapkan identitas pasien.

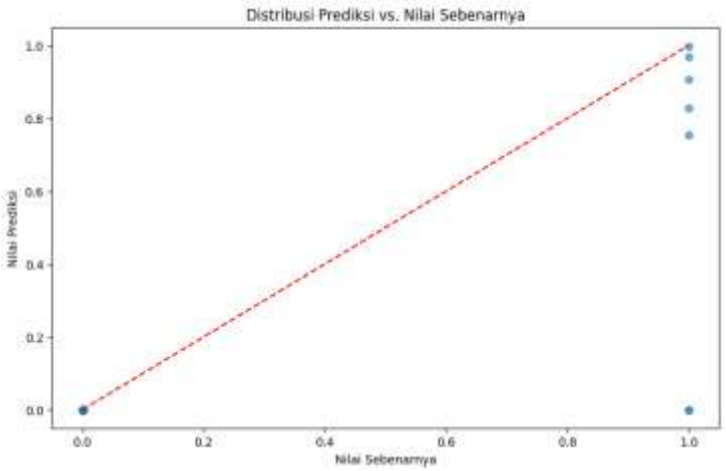
Tabel 1. Data Pasien

No.	Nama ART	Merokok	Batuk Berdahak >= 2 Minggu
1	LA UJU. SPDI	T	Y
2	LA ORE, SE	Y	Y
3	WA TELE	T	Y
4	LA ODE SABARUDIN	Y	T

3.2 Perbandingan Nilai Prediksi dan Nilai Aktual

Perbandingan nilai prediksi dan nilai aktual dilakukan untuk mengamati kesesuaian hasil keluaran model CatBoost dengan nilai sebenarnya pada data pengujian. Hasil perbandingan tersebut disajikan dalam bentuk grafik pada Gambar 2. Grafik ini digunakan untuk melihat pola perbedaan antara nilai prediksi dan nilai aktual.

Gambar 2. Grafik Perbandingan



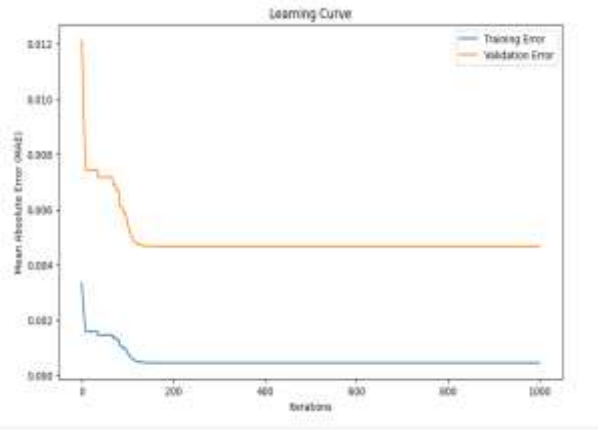
Garis diagonal dengan persamaan y = x menunjukkan kondisi ideal ketika nilai prediksi sama dengan nilai aktual. Titik-titik yang berada semakin dekat dengan garis tersebut menunjukkan selisih prediksi yang semakin kecil, sedangkan titik yang semakin jauh menunjukkan adanya perbedaan yang lebih besar.

Berdasarkan uraian grafik dalam naskah, sebagian titik pengamatan berada di sekitar garis diagonal, meskipun terdapat beberapa titik yang menunjukkan perbedaan antara nilai prediksi dan nilai aktual. Kondisi tersebut memperlihatkan bahwa kesesuaian prediksi perlu dinilai berdasarkan pola sebaran titik dan hasil pengukuran kesalahan model. Oleh karena itu, grafik perbandingan digunakan sebagai analisis visual yang melengkapi evaluasi kuantitatif pada tahap berikutnya.

3.3 Learning Curve

Learning curve digunakan untuk mengamati perubahan kinerja model selama proses pelatihan dengan membandingkan nilai kesalahan pada data pelatihan dan data validasi terhadap jumlah iterasi. Dalam penelitian ini, grafik learning curve ditampilkan pada Gambar 3 untuk memperlihatkan perubahan nilai Mean Absolute Error (MAE) selama proses pembelajaran model CatBoost.

Gambar 3. Grafik Perbandingan



Berdasarkan uraian pada naskah, kurva menunjukkan adanya perbedaan kinerja antara data pelatihan dan data validasi serta mengindikasikan kemungkinan terjadinya overfitting. Overfitting merupakan kondisi ketika model mempelajari pola pada data pelatihan dengan baik, tetapi kemampuan generalisasinya terhadap data yang tidak digunakan dalam pelatihan belum tentu sama baiknya.

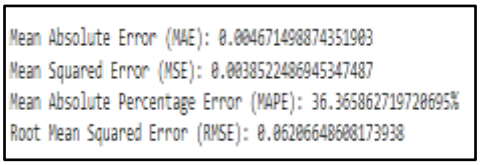
Indikasi tersebut perlu diperhatikan dalam mengevaluasi model, karena kinerja yang baik pada data pelatihan tidak secara otomatis menjamin kinerja yang sama pada data baru. Namun, tingkat overfitting tidak dapat ditentukan hanya berdasarkan uraian umum tanpa mengamati pola kurva secara langsung. Oleh sebab itu, interpretasi lebih lanjut perlu disesuaikan dengan nilai MAE pada data pelatihan dan validasi yang ditampilkan dalam grafik.

3.4 Hasil Evaluasi Model

Evaluasi dilakukan untuk mengukur perbedaan antara nilai prediksi model CatBoost dan nilai aktual pada data yang digunakan dalam pengujian. Penilaian kinerja model dalam penelitian ini menggunakan empat metrik, yaitu Mean Absolute Error (MAE), Mean Squared

Error (MSE), Mean Absolute Percentage Error (MAPE), dan Root Mean Squared Error (RMSE). Hasil perhitungan metrik tersebut disajikan pada Gambar 4.

Gambar 4. Hasil Evaluasi Model



Berdasarkan hasil evaluasi yang dilaporkan, model memperoleh nilai MAE sebesar 0,00467, MSE sebesar 0,00385, MAPE sebesar 36,37%, dan RMSE sebesar 0,062. Setiap metrik memberikan sudut pandang yang berbeda dalam menggambarkan kesalahan prediksi model.

Nilai MAE sebesar 0,00467 menunjukkan rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Nilai ini menggambarkan besarnya kesalahan absolut rata-rata berdasarkan skala nilai yang dievaluasi. Sementara itu, MSE sebesar 0,00385 mengukur rata-rata kuadrat kesalahan prediksi sehingga memberikan penalti lebih besar terhadap kesalahan yang nilainya tinggi.

Selanjutnya, MAPE sebesar 36,37% menunjukkan rata-rata kesalahan absolut dalam bentuk persentase berdasarkan perhitungan yang digunakan. Nilai ini perlu diperhatikan dalam interpretasi hasil karena persentase kesalahan dapat menjadi tidak stabil apabila terdapat nilai aktual nol atau mendekati nol. Dengan demikian, MAPE tidak dapat langsung diartikan sebagai tingkat akurasi model sebesar 63,63%.

Adapun nilai RMSE sebesar 0,062 merupakan akar kuadrat dari MSE dan menggambarkan besarnya kesalahan prediksi dengan memberikan bobot lebih besar terhadap kesalahan yang tinggi. Nilai RMSE perlu diinterpretasikan berdasarkan skala target yang digunakan dalam pemodelan.

Secara keseluruhan, keempat metrik tersebut memberikan gambaran kuantitatif mengenai kesalahan prediksi model CatBoost. Meskipun nilai MAE, MSE, dan RMSE yang dilaporkan relatif kecil secara numerik, penilaian bahwa model telah memiliki kinerja yang baik tetap memerlukan pertimbangan terhadap skala target, karakteristik data, serta kesesuaian metode evaluasi. Hasil tersebut juga perlu ditinjau bersama grafik perbandingan nilai prediksi dan nilai aktual serta learning curve agar interpretasi kinerja model tidak hanya didasarkan pada satu metrik.

4. Conclusion

Penelitian ini menerapkan algoritma CatBoost untuk memprediksi risiko penyakit tuberkulosis (TB) paru menggunakan 2.724 data pasien dari Puskesmas Kabangka, Kabupaten Muna. Kinerja model dievaluasi menggunakan empat metrik, yaitu Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), dan Root Mean Squared Error (RMSE). Hasil evaluasi yang dilaporkan menunjukkan nilai MAE sebesar 0,00467, MSE sebesar 0,00385, MAPE sebesar 36,37%, dan RMSE sebesar 0,062.

Hasil tersebut menggambarkan tingkat kesalahan prediksi berdasarkan metrik yang digunakan. Namun, interpretasi kinerja model perlu mempertimbangkan karakteristik variabel target dan kesesuaian metode evaluasi dengan jenis permasalahan yang diteliti. Oleh karena itu, hasil penelitian ini belum cukup untuk menyimpulkan kelayakan model sebagai alat skrining klinis tanpa evaluasi tambahan yang sesuai. Penelitian selanjutnya perlu memastikan

kesesuaian jenis model dengan variabel target serta menggunakan metrik evaluasi yang relevan agar kemampuan model dalam memprediksi risiko TB paru dapat dinilai secara lebih komprehensif.

5. References

Aditya, A., Umbara, F. R., & Sabrina, P. N. (2025). Klasifikasi Indeks Standar Pencemaran Udara Menggunakan Algoritma Catboost Dengan Teknik Balancing Data Random UnderSampling. *Jurnal Algoritma*, *22*(2). https://doi.org/10.33364/algoritma/v.22-2.2971

Syandika, N. D., & Yustanti, W. (2023). Deteksi Anomali Terhadap Pembatalan Transaksi Pada Platform Tiktok Shop dengan Algoritma Categorical Boosting (Catboost). *Journal of Informatics and Computer Science (JINACS)*, *5*(02), 149–156. https://doi.org/https://doi.org/10.48550/arXiv.1706.09516

Eko Fitra Firmandani, A. M., Hudawi AS, A., & Tholib, A. (2024). Optimasi Model CatBoost dengan Feature Selection dan Hyperparameter Tuning untuk Prediksi Nasabah Bank Potensial. *Academic Journal of Computer Science Research*, *6*(2). https://doi.org/10.38101/ajcsr.v6i2.15656

Ihzaniah, L. S., Setiawan, A., & Wijaya, R. W. N. (2023). Perbandingan Kinerja Metode Regresi K-Nearest Neighbor dan Metode Regresi Linear Berganda pada Data Boston Housing. *Jambura Journal of Probability and Statistics*, *4*(1). https://doi.org/10.34312/jjps.v4i1.18948

Kurniawan, A. C., & Salam, A. (2024). Seleksi Fitur Information Gain untuk Optimasi Klasifikasi Penyakit Tuberkulosis. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *8*(1). https://doi.org/10.30865/mib.v8i1.7122

Nurhaliza, N., Sabrina, P. N., & Ashaury, H. (2025). Klasifikasi Tuberkulosis (TBC) dengan Metode Random Forest Menggunakan Teknik Re-Sampling ADASYN-Tomek Links. *Jurnal Algoritma*, *22*(2). https://doi.org/10.33364/algoritma/v.22-2.2458

Nurul Abdillah, & Alfita Dewi. (2024). PENGEMBANGAN APLIKASI BERBASIS WEB DENGAN FRAMEWORK CODEIGNITER UNTUK PREDIKSI STATUS KEGAWATAN PASIEN TUBERCULOSIS MENGGUNAKAN ALGORITMA NBC DI KOTA PADANG. *The Indonesian Journal of Computer Science*, *13*(6). https://doi.org/10.33022/ijcs.v13i6.4455

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). Catboost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, *2018-December*.

Putri, E. R., & Arianto, D. B. (2024). Perbandingan Performa Algoritma Metode Bagging dan Boosting pada Prediksi Konsentrasi PM10 di Jakarta Utara. *Jurnal Nasional Teknologi Dan Sistem Informasi*, *10*(1). https://doi.org/10.25077/teknosi.v10i1.2024.72-81

Ramadhani, F., Iskandar Al-Idrus, S., Septiana, D., Arnita, A., & Retno Wahyuningrum, D. (2025). INTEGRASI MACHINE LEARNING DAN ANALISIS SPASIAL UNTUK PREDIKSI WILAYAH RAWAN TUBERKULOSIS DI PROVINSI SUMATERA UTARA. *Djtechno: Jurnal Teknologi Informasi*, *6*(2). https://doi.org/10.46576/djtechno.v6i2.6840

Syukur, M. A. A., Suhartono, S., & Chamidy, T. (2025). Prediksi Kualitas Udara Menggunakan Metode CatBoost. *JISKA (Jurnal Informatika Sunan Kalijaga)*, *10*(2). https://doi.org/10.14421/jiska.2025.10.2.249-258

Ula, M., Zulfikri, A., Ulva, A. F., & Rizal, R. A. (2023). Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru. *The Indonesian Journal of Computer Science*, *12*(1). https://doi.org/10.33022/ijcs.v12i1.3162