

PREDIKSI HASIL AKHIR PENGOBATAN PASIEN TUBERKULOSIS MENGUNAKAN ALGORITMA RANDOM FOREST

Jason Ikbal¹, Hardiansyah²

^{1,2} Prodi Studi Teknik Informatika, Universitas Pamulang, Jl. Puspitek, Buaran, Kec Pamulang,
Tangerang Selatan, Banten 15310, Indonesia

¹jasonikbal372@gmail.com, ²dosen02058@unpam.ac.id

Info Artikel

Riwayat Artikel:

Received July 16, 2026

Revised August 12, 2026

Accepted August 21, 2026

Corresponding Author:

Jason Ikbal

Email: Jasonikbal372@gmail.com



This is an open access
article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)
license.

Abstract – Tuberculosis (TB) remains one of the infectious diseases with a high number of cases in Indonesia. The lengthy treatment process causes treatment outcomes to be identified only after the completion of therapy, making early prediction essential to support clinical decision-making. This study aims to apply the Random Forest algorithm to predict tuberculosis treatment outcomes using historical patient data. The research adopts the Knowledge Discovery in Databases (KDD) methodology, consisting of data selection, preprocessing, transformation, data mining, and evaluation. The dataset consists of 11,067 patient records with 10 predictor attributes and one target attribute. Classification was performed using the Random Forest algorithm and evaluated through 10-fold cross-validation in Orange Data Mining. The experimental results achieved an accuracy of 88.3%, precision of 82.5%, recall of 82.2%, F1-score of 82.9%, and an Area Under the Curve (AUC) of 88.3%. Feature importance analysis revealed that TB diagnosis type, age, and healthcare facility type were the most influential attributes in predicting treatment outcomes. These findings indicate that the Random Forest algorithm can be utilized as a predictive model for tuberculosis treatment outcomes and has the potential to support decision-making in healthcare services.

Keywords: Data Mining, KDD, Prediction, Random Forest, Tuberculosis.

Abstrak – Tuberkulosis (TBC) masih menjadi salah satu penyakit menular dengan jumlah kasus yang tinggi di Indonesia. Lamanya proses pengobatan menyebabkan hasil akhir terapi baru diketahui setelah seluruh rangkaian pengobatan selesai, sehingga diperlukan pendekatan yang mampu membantu memprediksi hasil pengobatan sejak dini. Penelitian ini bertujuan menerapkan algoritma Random Forest untuk memprediksi hasil akhir pengobatan pasien Tuberkulosis berdasarkan data historis pasien. Metode yang digunakan adalah Knowledge Discovery in Databases (KDD) yang meliputi tahapan data selection, preprocessing, transformation, data mining, dan evaluation. Dataset yang digunakan terdiri atas 11.067 data pasien dengan 10 atribut prediktor dan 1 atribut target. Proses klasifikasi dilakukan menggunakan algoritma Random Forest dan dievaluasi menggunakan metode 10-Fold Cross Validation pada aplikasi Orange Data Mining. Hasil penelitian menunjukkan bahwa model memperoleh Accuracy sebesar 82.9%, Precision sebesar 82.5%, Recall sebesar 82.2%, F1-Score sebesar 82.9%, dan Area Under Curve (AUC) sebesar 88.3%. Analisis feature importance menunjukkan bahwa Tipe Diagnosis TBC, Umur, dan Jenis Fasilitas merupakan atribut yang paling berpengaruh terhadap hasil akhir pengobatan. Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu digunakan sebagai model prediksi hasil akhir pengobatan pasien Tuberkulosis serta berpotensi mendukung pengambilan keputusan dalam pelayanan kesehatan.

Kata Kunci: Data Mining, KDD, Prediksi, Random Forest, Tuberkulosis.

1. PENDAHULUAN

Tuberkulosis (TBC) merupakan salah satu penyakit menular yang masih menjadi permasalahan kesehatan masyarakat di dunia. Berdasarkan Global Tuberculosis Report, Indonesia termasuk negara

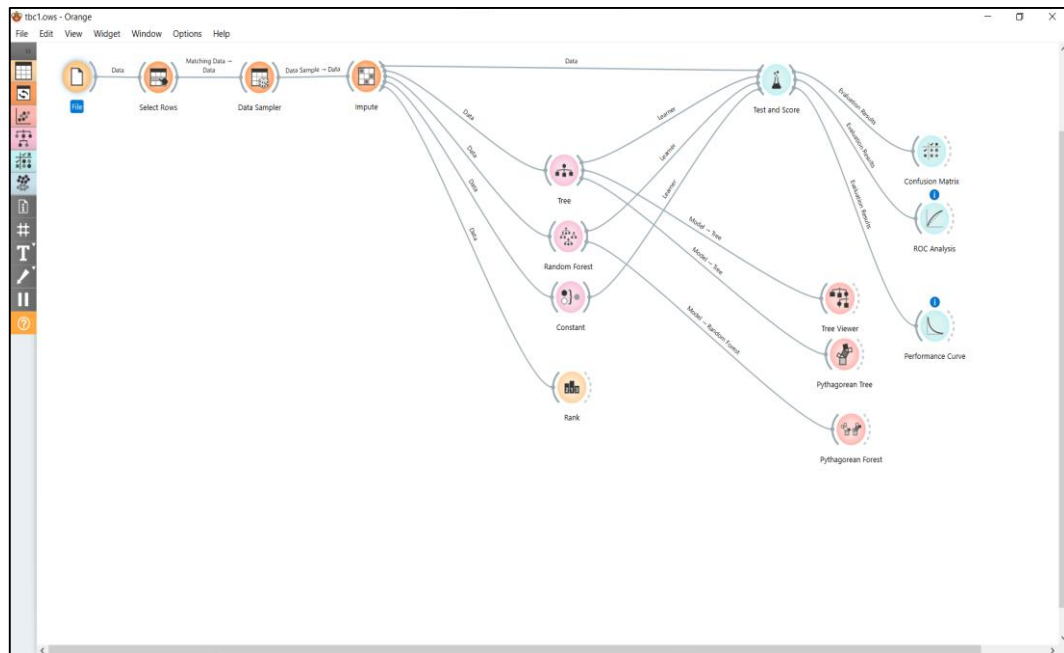
dengan beban Tuberkulosis tertinggi sehingga pengendalian penyakit ini masih menjadi prioritas nasional [1]. Kementerian Kesehatan Republik Indonesia juga menyatakan bahwa Tuberkulosis masih menjadi salah satu penyebab utama kesakitan dan kematian, sehingga diperlukan upaya yang mampu meningkatkan keberhasilan pengobatan pasien melalui pemanfaatan teknologi informasi dan analisis data [2]. Perkembangan Artificial Intelligence (AI), khususnya Machine Learning, memberikan peluang untuk memanfaatkan data historis pasien sebagai dasar dalam membangun sistem pendukung keputusan di bidang kesehatan. Salah satu pendekatan yang banyak digunakan adalah Data Mining, yaitu proses menggali pola dan informasi dari data berukuran besar sehingga menghasilkan pengetahuan yang dapat dimanfaatkan untuk pengambilan keputusan [3], [4]. Dalam penelitian ini digunakan kerangka kerja Knowledge Discovery in Databases (KDD) yang terdiri atas tahapan data selection, preprocessing, transformation, data mining, dan evaluation [3]. Salah satu algoritma klasifikasi yang banyak diterapkan pada permasalahan kesehatan adalah Random Forest. Algoritma ini merupakan metode ensemble learning yang membangun sejumlah decision tree melalui teknik bootstrap sampling dan feature randomness, kemudian menghasilkan prediksi akhir menggunakan mekanisme majority voting. Pendekatan tersebut membuat Random Forest memiliki kemampuan generalisasi yang baik, mampu menangani data numerik maupun kategorikal, serta lebih tahan terhadap overfitting dibandingkan algoritma pohon keputusan tunggal [5],[6][7].

Beberapa penelitian telah menerapkan algoritma machine learning pada kasus Tuberkulosis. Fahmi dan Sutisna [8] menerapkan algoritma Naïve Bayes untuk mengklasifikasikan gejala Tuberkulosis dan memperoleh hasil yang menunjukkan bahwa metode klasifikasi dapat membantu proses identifikasi penyakit. Ula et al. [9] memanfaatkan teknik machine learning untuk mengelompokkan tingkat risiko Tuberkulosis menggunakan K-Means dan Linear Regression. Harahap et al. [10] membandingkan algoritma Random Forest dan XGBoost pada klasifikasi penyakit paru-paru, sedangkan Tarigan dan Dahlan [11] mengoptimalkan pemilihan fitur menggunakan Forward Selection pada kasus penyakit paru-paru. Penelitian terbaru oleh Sugianto dan Triloka [12] menunjukkan bahwa Random Forest memiliki performa yang kompetitif dalam memprediksi hasil pengobatan Tuberkulosis di Indonesia. Meskipun demikian, penelitian yang memanfaatkan data historis pasien dari Dinas Kesehatan daerah untuk memprediksi hasil akhir pengobatan pasien Tuberkulosis masih relatif terbatas. Sebagian besar penelitian terdahulu berfokus pada diagnosis penyakit, klasifikasi gejala, atau perbandingan algoritma klasifikasi. Selain itu, belum banyak penelitian yang menganalisis atribut-atribut yang paling berpengaruh terhadap hasil akhir pengobatan menggunakan feature importance dari Random Forest. Berdasarkan permasalahan tersebut, penelitian ini bertujuan menerapkan algoritma Random Forest untuk memprediksi hasil akhir pengobatan pasien Tuberkulosis menggunakan data historis pasien dari Dinas Kesehatan Kota Tangerang Selatan. Dataset yang digunakan terdiri atas 11.067 data pasien dengan 10 atribut prediktor dan 1 atribut target. Model dikembangkan menggunakan metode Knowledge Discovery in Databases (KDD) dan dievaluasi menggunakan 10-Fold Cross Validation pada aplikasi Orange Data Mining [13].

Kontribusi penelitian ini adalah menghasilkan model prediksi hasil akhir pengobatan pasien Tuberkulosis berbasis Random Forest yang dapat dimanfaatkan sebagai pendukung pengambilan keputusan pada program pengendalian Tuberkulosis. Selain itu, penelitian ini juga menyajikan analisis feature importance untuk mengidentifikasi atribut yang paling memengaruhi hasil akhir pengobatan pasien. Kebaruan (novelty) penelitian terletak pada penerapan algoritma Random Forest terhadap data historis riil sebanyak 11.067 pasien dari Dinas Kesehatan Kota Tangerang Selatan untuk memprediksi enam kategori hasil akhir pengobatan serta mengidentifikasi faktor-faktor yang paling berpengaruh terhadap keberhasilan pengobatan.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan Knowledge Discovery in Databases (KDD) untuk membangun model prediksi hasil akhir pengobatan pasien Tuberkulosis menggunakan algoritma Random Forest. KDD merupakan proses sistematis dalam menemukan pengetahuan dari data yang meliputi tahapan data selection, preprocessing, transformation, data mining, dan evaluation [3], [4], [14]. Keseluruhan tahapan penelitian dilakukan menggunakan aplikasi Orange Data Mining yang menyediakan lingkungan visual untuk implementasi algoritma machine learning tanpa memerlukan penulisan kode program secara langsung [13].



Gbr. 1. Tahapan Penelitian Menggunakan Metode KDD

2.1 Dataset Penelitian

Dataset yang digunakan merupakan data historis pasien Tuberkulosis yang diperoleh dari Dinas Kesehatan Kota Tangerang Selatan. Dataset terdiri atas 11.067 data pasien yang berasal dari data tahun 2024 dan 2025. Variabel penelitian terdiri atas 10 atribut prediktor dan 1 atribut target, yaitu hasil akhir pengobatan pasien.

Tabel I. Variabel Penelitian

No	Variabel	Keterangan	Peran
1	Umur	Usia Pasien	Prediktor
2	Jenis kelamin	Jenis kelamin pasien	Prediktor
3	Fasyankes	Jenis Fasilitas Pelayanan Kesehatan	Prediktor
4	Tipe TBC	Tipe Diagnosis TBC	Prediktor
5	Riwayat Pengobatan	Riwayat Pengobatan Sebelumnya	Prediktor
6	Riwayat DM	Riwayat Diabetes Mellitus	Prediktor
7	Status HIV	Status Pemeriksaan HIV	Prediktor
8	TCM MTB Rif (Xpert)	Hasil Tes Cepat Molekuler TBC	Prediktor
9	DM	Status Diabetes Mellitus	Prediktor
10	HIV Hasil Tes	Hasil Pemeriksaan HIV	Prediktor
11	Hasil Akhir Pengobatan	Hasil Akhir Pengobatan Pasien TBC	Target

2.2 Data Selection

Tahap data selection bertujuan memilih atribut yang relevan terhadap tujuan penelitian. Pemilihan atribut dilakukan berdasarkan kajian literatur mengenai faktor-faktor yang memengaruhi hasil akhir pengobatan Tuberkulosis serta ketersediaan data pada dataset. Hasil tahap ini berupa sepuluh atribut prediktor dan satu atribut target yang digunakan dalam proses klasifikasi [3].

2.3 Data Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum proses pemodelan. Pada tahap ini dilakukan pemeriksaan terhadap missing value dan ketidaksesuaian data. Penanganan nilai yang hilang dilakukan menggunakan metode Average/Most Frequent melalui widget Impute pada Orange Data Mining. Pendekatan ini mengisi atribut numerik menggunakan nilai rata-rata (average) dan atribut

kategorikal menggunakan nilai yang paling sering muncul (most frequent), sehingga seluruh data dapat digunakan dalam proses pelatihan model .

2.4 Data Transformation

Tahap transformasi dilakukan untuk menyesuaikan struktur data agar dapat diproses oleh algoritma Random Forest. Seluruh atribut dikonversi ke format yang sesuai dengan kebutuhan proses klasifikasi sehingga data siap digunakan pada tahap pembentukan model [3], [14].

2.5 Data Mining

Proses data mining dilakukan menggunakan algoritma Random Forest. Algoritma ini merupakan metode ensemble learning yang membangun sejumlah Decision Tree melalui teknik bootstrap sampling dan pemilihan atribut secara acak (feature randomness). Hasil prediksi setiap Decision Tree kemudian digabungkan menggunakan mekanisme majority voting sehingga menghasilkan model yang lebih stabil dan memiliki kemampuan generalisasi yang baik dibandingkan pohon keputusan tunggal [5], [7], [15].

Implementasi model dilakukan menggunakan widget Random Forest pada Orange Data Mining. Selanjutnya model dievaluasi menggunakan metode 10-Fold Cross Validation, yaitu teknik validasi yang membagi dataset menjadi sepuluh bagian sehingga setiap bagian memperoleh kesempatan menjadi data latih maupun data uji secara bergantian. Teknik ini banyak digunakan untuk mengurangi bias evaluasi dan memperoleh estimasi performa model yang lebih representatif [13].

2.6 Evaluation

Tahap evaluasi bertujuan mengukur kemampuan model dalam memprediksi hasil akhir pengobatan pasien Tuberkulosis. Evaluasi dilakukan menggunakan widget Test and Score pada Orange Data Mining dengan metode 10-Fold Cross Validation. Kinerja model diukur menggunakan metrik Accuracy, Precision, Recall, F1-Score, dan Area Under Curve (AUC). Selain itu, dilakukan analisis Feature Importance menggunakan widget Rank untuk mengidentifikasi atribut yang paling berpengaruh terhadap hasil prediksi [5], [13] .

3. HASIL DAN PEMBAHASAN

3.1 Hasil Evaluasi Model

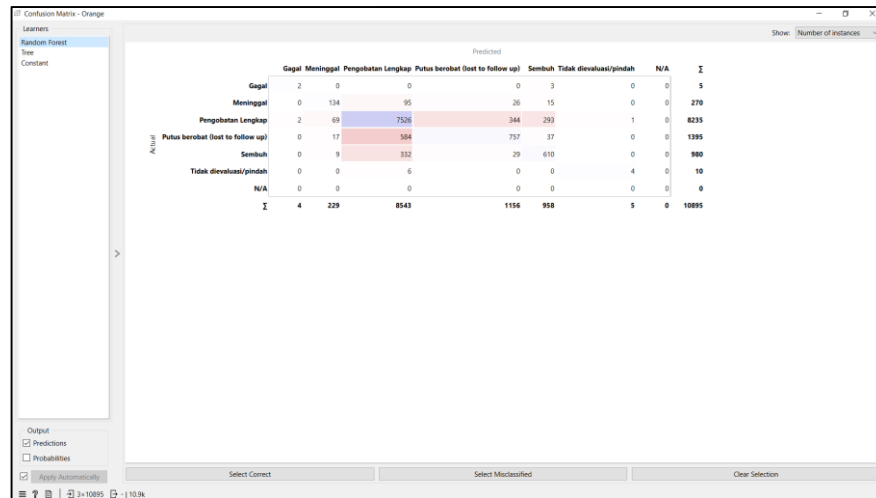
Model prediksi hasil akhir pengobatan pasien Tuberkulosis dibangun menggunakan algoritma Random Forest dengan metode evaluasi 10-Fold Cross Validation pada aplikasi Orange Data Mining. Penggunaan teknik validasi ini bertujuan memperoleh estimasi performa model yang lebih objektif karena setiap data memperoleh kesempatan menjadi data latih dan data uji secara bergantian [13].

Kinerja model diukur menggunakan lima metrik evaluasi, yaitu Accuracy, Precision, Recall, F1-Score, dan Area Under Curve (AUC). Hasil pengujian ditunjukkan pada Tabel II.

Tabel II. Hasil Evaluasi Model Random Forest

Parameter	Nilai
AUC	0.883
CA	0.829
Precision	0.825
Recall	0.822
F1-Score	0.829

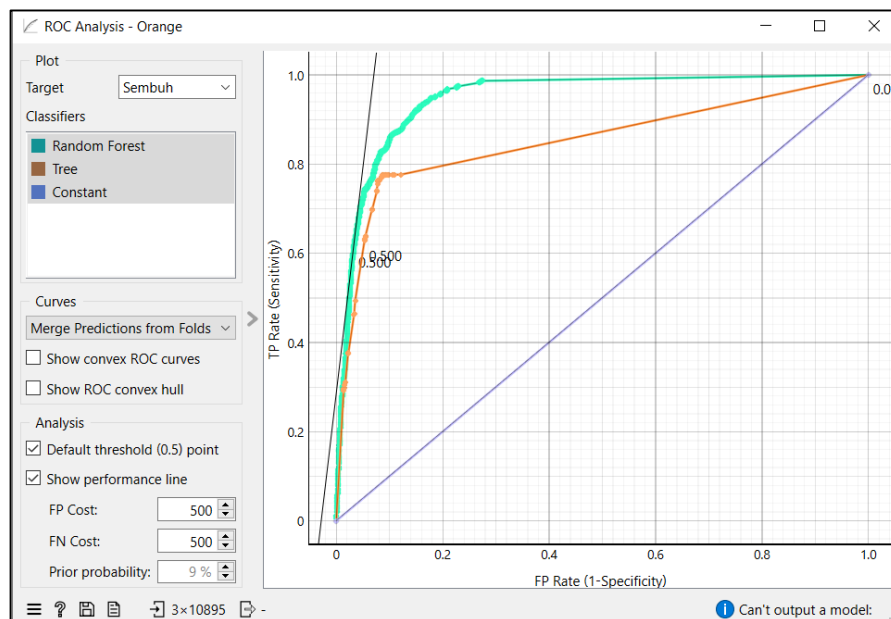
Berdasarkan Tabel II, model Random Forest memperoleh Accuracy sebesar 82.9%, yang menunjukkan bahwa sebagian besar data pasien berhasil diklasifikasikan dengan benar. Nilai Precision sebesar 82.5% menunjukkan bahwa prediksi yang dihasilkan model memiliki tingkat ketepatan yang cukup baik, sedangkan Recall sebesar 82.2% menunjukkan kemampuan model dalam mengenali data pada setiap kategori hasil pengobatan. Nilai F1-Score sebesar 82.9% menggambarkan keseimbangan antara Precision dan Recall, sementara AUC sebesar 88.3% menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan kelas hasil akhir pengobatan [5]. Untuk mengetahui distribusi hasil klasifikasi pada masing-masing kelas, dilakukan analisis menggunakan Confusion Matrix.



Gbr. 2. Confusion Matrix Random Forest

Berdasarkan Confusion Matrix, sebagian besar data berhasil diprediksi pada kelas yang benar. Namun demikian, masih terdapat kesalahan klasifikasi pada beberapa kelas dengan jumlah data yang relatif sedikit, seperti kategori Gagal, Meninggal, dan Tidak Dievaluasi/Pindah. Ketidakseimbangan jumlah data antar kelas (class imbalance) menjadi salah satu faktor yang memengaruhi performa model sehingga beberapa kelas minoritas lebih sulit diprediksi secara akurat.

Selanjutnya dilakukan analisis menggunakan ROC Curve untuk mengevaluasi kemampuan model dalam membedakan setiap kelas hasil pengobatan.

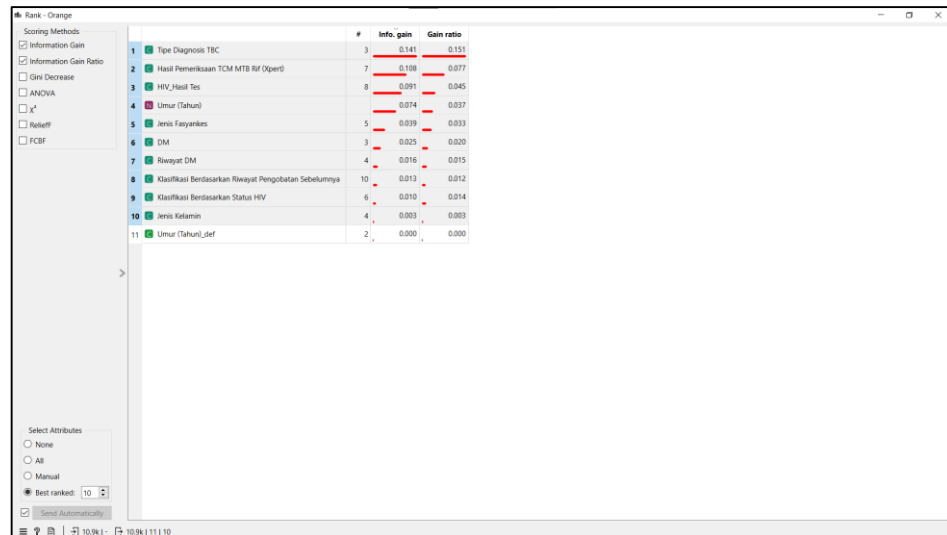


Gbr. 3. ROC Curve Random Forest

Nilai AUC sebesar 68,6% menunjukkan bahwa model memiliki kemampuan diskriminasi yang cukup baik dalam membedakan kelas hasil akhir pengobatan. Nilai tersebut mengindikasikan bahwa model mampu memberikan prediksi yang lebih baik dibandingkan klasifikasi secara acak, meskipun masih terdapat ruang untuk peningkatan performa melalui optimasi parameter maupun penambahan atribut prediktor.

3.2 Analisis Feature Importance

Salah satu keunggulan algoritma Random Forest adalah kemampuannya menghasilkan nilai feature importance, yaitu tingkat kontribusi masing-masing atribut terhadap proses klasifikasi [5].



Gbr. 4. Feature Importance Random Forest

Hasil analisis menunjukkan bahwa Tipe Diagnosis TBC, Umur, dan Jenis Fasyankes merupakan tiga atribut dengan tingkat kepentingan tertinggi. Tipe Diagnosis TBC menjadi atribut paling dominan karena berkaitan dengan metode penegakan diagnosis pasien, sedangkan Umur berpengaruh terhadap kondisi fisiologis dan respons pasien terhadap terapi. Selain itu, Jenis Fasyankes menunjukkan bahwa karakteristik fasilitas pelayanan kesehatan turut memengaruhi keberhasilan pengobatan pasien.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma Random Forest mampu membangun model prediksi hasil akhir pengobatan pasien Tuberkulosis dengan performa yang cukup baik. Kemampuan tersebut didukung oleh karakteristik Random Forest sebagai metode ensemble learning yang menggabungkan sejumlah Decision Tree sehingga mampu meningkatkan stabilitas model dan mengurangi risiko overfitting [5], [7]. Temuan penelitian ini sejalan dengan penelitian Harahap et al. [10] yang menunjukkan bahwa Random Forest memberikan performa klasifikasi yang kompetitif dibandingkan algoritma lain pada kasus penyakit paru-paru. Penelitian Sugianto dan Triloka [12] juga menunjukkan bahwa Random Forest efektif digunakan untuk memprediksi hasil pengobatan Tuberkulosis berdasarkan data pasien di Indonesia. Selain itu, penelitian Fahmi dan Sutisna [8] membuktikan bahwa teknik klasifikasi mampu membantu proses identifikasi kasus Tuberkulosis menggunakan data kesehatan. Sementara itu, penelitian Tarigan dan Dahlan [11] menunjukkan bahwa optimasi pemilihan fitur dapat meningkatkan performa model klasifikasi. Temuan tersebut mendukung hasil penelitian ini, di mana analisis feature importance berhasil mengidentifikasi atribut-atribut yang memiliki kontribusi terbesar terhadap hasil prediksi.

Berbeda dengan penelitian Ula et al. [9] yang menggunakan pendekatan clustering untuk menentukan tingkat risiko Tuberkulosis, penelitian ini berfokus pada prediksi hasil akhir pengobatan menggunakan pendekatan klasifikasi berbasis Random Forest. Selain menghasilkan model prediksi, penelitian ini juga memberikan informasi mengenai faktor-faktor yang paling berpengaruh terhadap keberhasilan pengobatan pasien melalui analisis feature importance. Secara praktis, model yang dihasilkan dapat dimanfaatkan sebagai sistem pendukung keputusan bagi tenaga kesehatan dalam mengidentifikasi pasien yang berpotensi mengalami kegagalan pengobatan atau putus berobat sejak tahap awal terapi. Dengan demikian, intervensi dapat dilakukan lebih cepat untuk meningkatkan keberhasilan pengobatan dan mendukung program pengendalian Tuberkulosis.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma Random Forest untuk memprediksi hasil akhir pengobatan pasien Tuberkulosis berdasarkan data historis pasien dari Dinas Kesehatan Kota Tangerang Selatan menggunakan metode Knowledge Discovery in Databases (KDD). Berdasarkan hasil evaluasi menggunakan metode 10-Fold Cross Validation, model yang dibangun memperoleh nilai Accuracy sebesar 82.9%, Precision sebesar 82.5%, Recall sebesar 82.2%, F1-Score sebesar 82.9%, dan Area Under Curve (AUC) sebesar 88.3%, sehingga menunjukkan bahwa Random Forest memiliki kemampuan yang cukup baik dalam memprediksi hasil akhir pengobatan pasien Tuberkulosis. Analisis feature importance menunjukkan bahwa Tipe Diagnosis TBC, Umur, dan Jenis Fasyankes merupakan atribut yang paling berpengaruh terhadap hasil prediksi. Temuan ini menunjukkan bahwa pemanfaatan data historis pasien melalui algoritma machine learning dapat mendukung pengambilan keputusan pada program pengendalian Tuberkulosis, khususnya dalam mengidentifikasi pasien yang berpotensi mengalami hasil pengobatan yang kurang optimal sejak tahap awal terapi. Untuk penelitian selanjutnya, model prediksi dapat dikembangkan melalui optimasi hiperparameter, penambahan variabel klinis maupun faktor sosial yang lebih lengkap, serta perbandingan dengan algoritma ensemble learning lainnya seperti XGBoost, LightGBM, dan CatBoost untuk memperoleh performa prediksi yang lebih baik.

DAFTAR PUSTAKA

- [1] W. H. Organization, *Global tuberculosis report 2023*. Geneva: World Health Organization, 2023. [Online]. Available: <https://www.who.int/publications/i/item/9789240083851>
- [2] Kementerian Kesehatan, "Tuberculosis." Accessed: Jun. 02, 2026. [Online]. Available: <https://ayosehat.kemkes.go.id/topik-penyakit/kesehatan-lainnya/tuberkulosis?>
- [3] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd Editio. Waltham, MA: Morgan Kaufmann Publishers (Elsevier), 2012. [Online]. Available: <https://shop.elsevier.com/books/data-mining-concepts-and-techniques/han/978-0-12-381479-1>
- [4] Aggarwal and C. C, Aggarwal, C. C. (2015). *An introduction to data mining*. In *Data mining: The textbook* (pp. 1-26). Cham: Springer International Publishing. Cham: Springer International Publishing, 2015. doi: 10.1007/978-3-319-14142-8
- [5] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.
- [6] Y. Heryadi and T. Wahyuno, *Machine Learning: Konsep dan Implementasi*, 1st ed. Yogyakarta: penerbit gava media, 2020.
- [7] C. P. López, *Data Mining and Machine Learning: Predictive Techniques: Ensemble Methods, Boosting, Bagging, Random Forest, Decision Trees and Regression Trees*. lulu.com, 2021.
- [8] H. Fahmi and Sutisna, "Implementasi Data Mining Klasifikasi Gejala Penyakit TB Menggunakan Algoritma Naive Bayes pada Studi Kasus Puskesmas Pegangsaan Dua B," *J. Indones. Manaj. Inform. dan Komun.*, vol. 5, no. 3, pp. 2888–2898, Sep. 2024. doi: 10.35870/jimik.v5i3.970
- [9] M. Ula, A. Zulfikri, A. Faridhatul Ulva, and R. Achmad Rizal, "Penerapan Machine Learning Clustering K-Means dan Linear Regression dalam Penentuan Tingkat Resiko Tuberkulosis Paru," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 336–348, Feb. 2023. doi: 10.33022/ijcs.v12i1.3162
- [10] R. Harahap, M. Irpan, M. Azzuhri Dinata, and L. Efrizoni, "PERBANDINGAN ALGORITMA RANDOM FOREST DAN XGBOOST UNTUK KLASIFIKASI PENYAKIT PARU-PARU BERDASARKAN DATA DEMOGRAFI PASIEN," Aug. 2024. doi: 10.36050/3v3xwn06
- [11] L. Rangga Aditya Tarigan and Dahlan, "OPTIMALISASI FITUR DENGAN FORWARD SELECTION PADA ESTIMASI TINGKAT PENYAKIT PARU-PARU MENGGUNAKAN ALGORITMA," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 5, pp. 10341–10348, Oct. 2024. doi: 10.36040/jati.v8i5.11064
- [12] D. Sugianto and J. Triloka, "Prediction of Tuberculosis Treatment Outcomes in Indonesia Using Support Vector Machine and Random Forest," Oct. 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>. doi: 10.30871/jaic.v9i5.10018
- [13] J. Demšar *et al.*, "Orange: Data mining toolbox in python," *J. Mach. Learn. Res.*, vol. 14, pp. 2349–2353, 2013. doi: 10.5555/2567709.2567736
- [14] B. Santoso and A. Umam, *Data mining dan Big Data Analytics : Teori dan Implementasi Menggunakan Python & Apache Spark*, 2nd ed. Yogyakarta: Penebar Media Pustaka, 2018.
- [15] R. Toscano, "Machine Learning," 2024, pp. 203–238. doi: 10.1007/978-3-031-52459-2_7.