

yang sebelumnya 768 data menjadi 760 data setelah dilakukan tahapan *pre-processing*. *Accuracy*, *Recall*, *Precision*, *F1 Score*, yang didasarkan pada Confusion Matrix, dan grafik ROC/AUC digunakan untuk mengevaluasi hasil penelitian. Penerapan algoritma *machine learning* dalam prediksi penyakit diabetes berhasil dilakukan dengan baik. Berdasarkan hasil prediksi analisis *accuracy* penyakit diabetes menggunakan algoritma *Decision Tree* dan *Random Forest* didapatkan hasil perbandingan terbaik untuk menentukan metode mana yang lebih efisien dalam memprediksi kasus ini. Pada proses perhitungan K-fold 10 dan SMOTE, penerapan algoritma *Random Forest* menggunakan SMOTE sangat efisien dibandingkan dengan DC + SMOTE yang memiliki tingkat prediksi kurang akurat. Evaluasi model juga berhasil dilakukan dengan nilai prediksi yang sangat baik sebesar 88,9% untuk *Random Forest* dengan nilai akurasi tertinggi pada kasus penyakit Diabetes. Dengan hasil Kurva ROC dengan nilai *Area Under the Curve (AUC)* 89,0%.

DAFTAR PUSTAKA

- [1] M. S. Efendi and H. A. Wibawa, "<i>Prediksi Penyakit Diabetes Menggunakan Algoritma ID3 dengan Pemilihan Atribut Terbaik (Diabetes Prediction using ID3 Algorithm with Best Attribute Selection) </i>," <i>Juita</i>, vol. VI, no. 1, pp. 29–35, 2018.
- [2] B. M. K. P, S. P. R, N. R K, and A. K, "Type 2: Diabetes mellitus prediction using Deep Neural Networks classifier," *Int. J. Cogn. Comput. Eng.*, vol. 1, no. July, pp. 55–61, 2020, doi: 10.1016/j.ijcce.2020.10.002.
- [3] A. S. Putri, "Tahun 2018 Penderita Diabetes di Indonesia Meningkat," 2018, 2018. <https://www.fimela.com/lifestyle/read/3739252/tahun-2018-penderita-diabetes-di-indonesia-meningkat>
- [4] R. Pahlevi, "Kasus Kematian Akibat Diabetes di Indonesia Terbesar Keenam di Dunia," 2021, 2021. <https://databoks.katadata.co.id/datapublish/2021/11/26/kasus-kematian-akibat-diabetes-di-indonesia-terbesar-keenam-di-dunia>
- [5] J. J. Khanam and S. Y. Foo, "A comparison of machine learning algorithms for diabetes prediction," *ICT Express*, vol. 7, no. 4, pp. 432–439, 2021, doi: 10.1016/j.icte.2021.02.004.
- [6] D. Care and S. S. Suppl, "1. Improving care and promoting health in populations: Standards of medical care in diabetes-2020," *Diabetes Care*, vol. 43, no. January, pp. S7–S13, 2020, doi: 10.2337/dc20-S001.
- [7] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM Resampling Technique for Handling Class Imbalance in the Classification of Diabetes using C4.5, Random Forest, and SVM," *Agustus*, vol. 20, no. 3, pp. 352–361, 2021, [Online]. Available: <https://www.kaggle.com/uciml/pima-indians-diabetes-database>.
- [8] G. Swapna, R. Vinayakumar, and K. P. Soman, "Diabetes detection using deep learning algorithms," *ICT Express*, vol. 4, no. 4, pp. 243–246, 2018, doi: 10.1016/j.icte.2018.10.005.
- [9] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>
- [10] J. B. Junior, R. R. Saedudin, and V. P. Widharta, "Perbandingan Akurasi Algoritma Decision Tree Dan Algoritma Support Vector Machine Pada Penyakit Diabetes," vol. 8, no. 5, pp. 9749–9756, 2021.
- [11] R. S. R. io, "Pima Indians Diabetes Database," *Subgroup Discovery dan Bump Hunting*, 2020.
- [12] R. Siringoringo, "KLASIFIKASI DATA TIDAK SEIMBANG MENGGUNAKAN ALGORITMA SMOTE DAN k-NEAREST NEIGHBOR," *Isd*, vol. 3, no. 1, pp. 2528–5114, 2018.
- [13] N. Azizah, "Komparasi Metode Klasifikasi Decision Tree Algoritma C4.5 Dan Random Forest Untuk Prediksi Penyakit Stroke," 2021.
- [14] A. Franseda, W. Kurniawan, S. Anggraeni, and W. Gata, "Integrasi Metode Decision Tree dan SMOTE untuk Klasifikasi Data Kecelakaan Lalu Lintas," *J. Sist. dan Teknol. Inf.*, vol. 8, no. 3, p. 282, 2020, doi: 10.26418/justin.v8i3.40982.
- [15] A. Primajaya and B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 27, 2018, doi: 10.24014/ijaidm.v1i1.4903.

- [16] dqlab.id, "Studi Kasus Random Forest Machine Learning untuk Pemula Data."
- [17] A. Samosir, M. Hasibuan, W. E. Justino, and T. Hariyono, "Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung," pp. 214–222, 2021.
- [18] N. Khasanah, R. Komarudin, N. Afni, Y. I. Maulana, and A. Salim, "Skin Cancer Classification Using Random Forest Algorithm," *Sisfotenika*, vol. 11, no. 2, p. 137, 2021, doi: 10.30700/jst.v11i2.1122.
- [19] J. J. Pangaribuan, C. Tedja, and S. Wibowo, "PERBANDINGAN METODE ALGORITMA C4.5 DAN EXTREME LEARNING MACHINE UNTUK MENDIAGNOSIS PENYAKIT JANTUNG KORONER," 2019.
- [20] E. Prasetyo, B. Prasetyo, and P. Korespondensi, "PENINGKATAN AKURASI KLASIFIKASI ALGORITMA C4.5 MENGGUNAKAN TEKNIK BAGGING PADA DIAGNOSIS PENYAKIT JANTUNG INCREASED CLASSIFICATION ACCURACY C4.5 ALGORITHM USING BAGGING TECHNIQUES IN DIAGNOSING HEART DISEASE," vol. 7, no. 5, pp. 1035–1040, 2020, doi: 10.25126/jtiik.202072379.

