JURNAL SIMPUL INOVASI
Journal of Innovation – Hub
Pusat Inkubator Bisnis – Universitas Jenderal Soedirman
E-ISSN: 3109-189x
Volume (2), Issue (2), Halaman 75-80, Desember 2025
Pusat Inkubator Bisnis Unsoed

Evaluasi Strategi Penanganan *Imbalancedd Class* Terhadap Akurasi Algoritma *Machine Learning* (ML): Studi Eksperimen Pima Indians Diabetes *Dataset*

Agus Wantoro[1*], Nur Aminudin[2], Agus Irawan[3], Dwi Feriyanto[4]
[1,2,4]Teknik Informatika, Fakultas Teknologi dan Informatika, Universitas Aisyah Pringsewu
Pringsewu, Lampung, Indonesia
[3]Sistem Informasi, Fakultas Teknologi dan Ilmu Komputer, Institut Bakri Nusantara
Pringsewu, Lampung, Indonesia
*E-mail: aguswantoro@aisyahuniversity.ac.id (*corresponding author*)

Abstrak
Ketidakseimbangan kelas merupakan salah satu tantangan utama dalam pengembangan model klasifikasi untuk prediksi penyakit, termasuk diabetes. Pima Indians Diabetes Dataset (PIDD) sering digunakan sebagai basis evaluasi model diagnosis diabetes, memiliki distribusi data yang tidak seimbang antara kelas positif (diabetes) dan negatif (non-diabetes). Hal ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas dan gagal mengenali pola pada kelas minoritas yang terkadang justru lebih kritis. Penelitian ini bertujuan untuk mengevaluasi pengaruh berbagai strategi penanganan *imbalancedd class* menggunakan metode *Synthetic Minority Oversampling Technique* (SMOTE) terhadap kinerja beberapa algoritma klasifikasi seperti Naive Bayes, AdaBoost, J48, Random Tree, Random Forest, dan Support Vector Machine (SVM). Evaluasi dilakukan berdasarkan akurasi dan waktu komputasi. Hasil eksperimen menunjukkan bahwa algoritma Random Forest menghasilkan performa terbaik dalam mendeteksi kasus diabetes, dengan peningkatan signifikan pada nilai akurasi dibandingkan model yang dilatih tanpa penyeimbangan kelas. Temuan ini mengindikasikan bahwa pemilihan strategi penanganan *imbalanced class* yang tepat dapat secara signifikan meningkatkan kemampuan model dalam klasifikasi medis, khususnya untuk data yang tidak seimbang.

Kata kunci: *Imbalanced class*, Machine Learning, Klasifikasi, Diabetes

1 Pendahuluan

Diabetes merupakan salah satu masalah kesehatan global yang terus meningkat prevalensinya [1]. Deteksi dini dan diagnosis yang akurat sangat penting untuk mencegah komplikasi serius yang dapat ditimbulkan oleh penyakit ini [2]. Dalam beberapa tahun terakhir, pendekatan berbasis *Machine Learning* (ML) telah banyak digunakan untuk membantu dalam proses diagnosis diabetes [3]. Namun, tantangan signifikan yang sering dihadapi dalam penerapan model klasifikasi ini adalah ketidakseimbangan kelas dalam dataset, di mana jumlah data pada kelas mayoritas "non-diabetes" lebih banyak dibandingkan dengan kelas minoritas "diabetes" [4]

Ketidakseimbangan kelas ini dapat menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas, sehingga mengurangi kemampuan model dalam mendeteksi kasus pada kelas minoritas [5]. Beberapa teknik telah dikembangkan untuk mengatasi masalah ini, seperti *Synthetic Minority Over-sampling Technique* (SMOTE) [3][6]. Teknik-teknik ini bertujuan untuk menyeimbangkan distribusi kelas dalam dataset sebelum proses pelatihan model dilakukan [6].

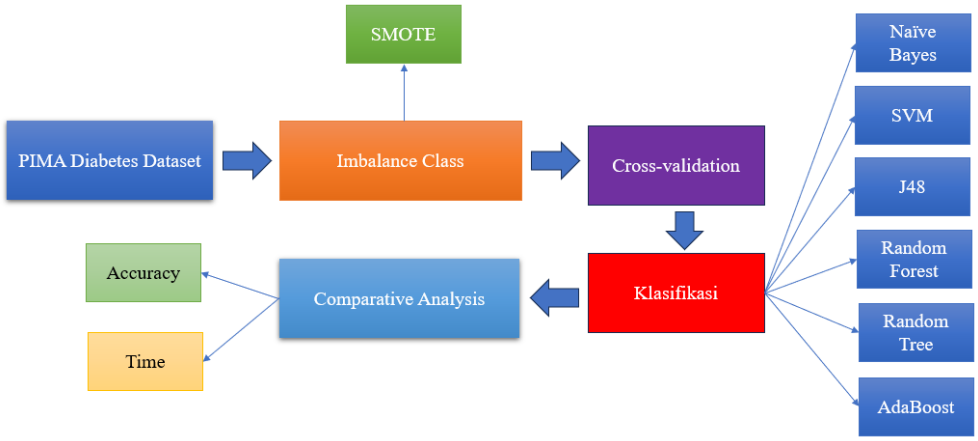
Pima Indians Diabetes Dataset (PIDD) merupakan salah satu *dataset* yang sering digunakan dalam penelitian terkait prediksi diabetes [7]. Namun, *dataset* ini juga mengalami masalah ketidakseimbangan kelas yang signifikan. Beberapa studi telah menunjukkan bahwa penerapan teknik penanganan ketidakseimbangan kelas dapat meningkatkan kinerja model klasifikasi pada dataset ini [8]. Misalnya, penelitian oleh [9] menunjukkan bahwa penggunaan teknik *Oversampling* dapat meningkatkan akurasi model dalam mendeteksi kasus diabetes [5].

Meskipun berbagai teknik telah dikembangkan dan diterapkan, masih terdapat kebutuhan untuk melakukan evaluasi komprehensif terhadap efektivitas masing-masing teknik dalam konteks prediksi diabetes. Evaluasi ini penting untuk menentukan strategi terbaik dalam menangani ketidakseimbangan kelas, sehingga dapat meningkatkan akurasi dan reliabilitas model klasifikasi dalam mendeteksi diabetes.

2 Metodologi

Bagian ini menguraikan prosedur dan tahapan penelitian. Tahap penelitian diawali dengan pengumpulan data. Sebelum data dimasukkan ke dalam model, perlu dilakukan *pra*-pemrosesan data. Pada tahap ini, data dibersihkan dan disesuaikan untuk diproses ke langkah berikutnya melakukan keseimbangan kelas menggunakan teknik SMOTE. Tahap

berikutnya adalah pemilihan jumlah *k-fold validation* untuk klasifikasi metode ML seperti Naive Bayes, AdaBoost, J48, Random Tree, Random Forest, dan Support Vector Machine (SVM). Desain kerangka penelitian diilustrasikan pada Gambar 1.



Gambar 1. Desain kerangka penelitian

2.1 Dataset

Dataset adalah kumpulan data terstruktur yang disusun untuk analisis, penelitian, atau pelatihan model ML. Dataset umumnya disajikan dalam bentuk tabel (dengan baris dan kolom) dan digunakan untuk mengidentifikasi pola, membangun model, dan mendapatkan wawasan dari data [10]. Penelitian ini menggunakan *dataset* PIDD yang diambil dari *www.kaggle.com/uciml/pima-indians-diabetes-database*. *Dataset* memiliki 768 record, 8 fitur, dan 2 class. Jumlah data setiap class ditampilkan pada Tabel 1.

Tabel 1. Jumlah Class

Diabetes	**Non-diabetes**
268	500

Tabel 1 menampilkan jumlah *class "diabates"* dan "*non-diabetes"* tidak seimbang. Jumlah *class* "non-diabetes" mayoritas dan *class* "diabetes" minoritas. Hal ini mengakibatkan akurasi klasifikasi algoritma ML tidak optimal karena cenderung pada *class* mayoritas [11]. Fitur-fitur yang digunakan dalam dataset PIDD disajikan dalam Tabel 2.

Tabel 2. Fitur dan attribute dataset

Kode	**Fitur**	**Keterangan**
f1	*Pregnancies*	Jumlah pregnancies
f2	*Glucose Plasma*	Kadar glukosa plasma dua jam setelah mengonsumsi glukosa
f3	*Blood Pressure*	Diastolic blood pressure (mm Hg)
f4	*Skin*	Ketebalan lipatan kulit pada trisep lengan atas (mm)
f5	*Insulin*	Kadar serum insulin dalam darah dua jam setelah tes glukosa (lh/ml)
f6	*BMI*	Body mass index kg/(Height in m)], indeks yang digunakan untuk mengevaluasi berat relatif seseorang
f7	*Pedigree*	Fungsi silsilah diabetes adalah nilai yang mengukur faktor risiko genetik berdasarkan riwayat keluarga diabetes.
f8	*Age*	Usia pasien
f9	*Class*	"diabetes" dan "non-diabetes"

2.2 SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) adalah pengembangan dari metode *oversampling,* dimana cara kerja metode ini adalah dengan membangkitkan sampel baru yang berasal dari kelas minoritas untuk membuat proporsi data menjadi lebih seimbang dengan cara sampling ulang sampel kelas minoritas [3], [6]

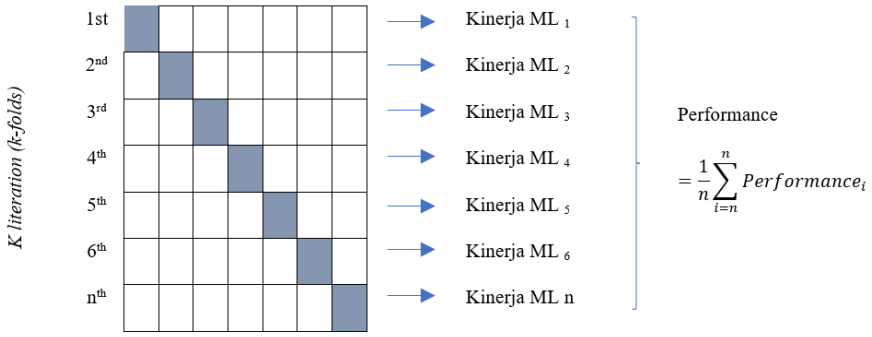
2.3 Performance Measurement (PM)

PM digunakan untuk mengukur kinerja algoritma ML menggunakan perhitungan jumlah prediksi tepat yang dibagi seluruh jumlah data [8][12]. Perhitungan akurasi kinerja algoritma ML menggunakan persamaan (1)

$$Accuracy = \frac{Semua\ perediksi\ benar}{Semua\ data} \quad (1)$$

2.4 Cross-validation Classification

Klasifikasi merupakan salah satu bentuk teknik penambangan data yang sedang populer saat ini. Strategi ini menggunakan berbagai metode untuk menilai data yang tersedia guna menghasilkan prediksi diabetes [13]. Model klasifikasi akan divalidasi menggunakan *k-fold cross-validation*. Metode *cross-validation* umumnya digunakan untuk training set [14]. Cross-validation digunakan untuk melakukan pengujian algoritma ML pada Tabel 3 dan 4. Gambar 2 menampilkan jumlah *k-fold cross-validation*



Gambar 2. *Prosedur k-fold validation*

3 Hasil dan pembahasan

Kami menggunakan *software* WEKA (versi 3.9.2). Platform ini menyederhanakan konstruksi beberapa data teknik analisis. WEKA memiliki kemampuan mengategorikan, melakukan regresi, klasifikasi, menghilangkan fitur, membuat aturan asosiasi, dan penyesuaian class pada dataset [1]. Kami melakukan penyeimbangan class menggunakan pendekatan *oversampling class* menggunakan metode SMOTE dengan *k-fold*=10 karena temuan kami pilihan ini terbaik terhadap klasifikasi algoritma ML. Hasil perbedaan jumlah *class* setelah dilakukan keseimbangan *class* dalam Tabel 3.

Tabel 3. Jumlah Class Awal dan SMOTE

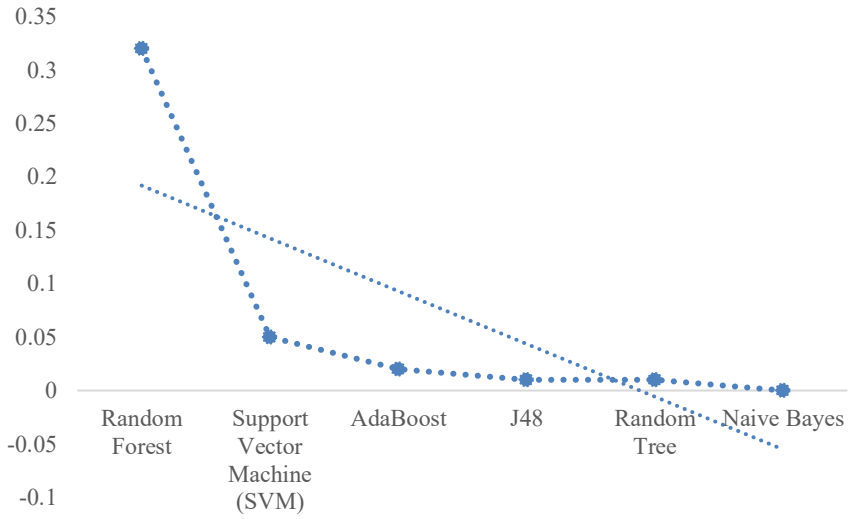
Class	**Awal**	**SMOTE**
Diabetes	268	530
Non	500	500
Total	**768**	**1030**

Tabel 3 menunjukkan bahwa penerapan metode SMOTE meningkatkan jumlah class dan record. Selanjutnya kami melakukan pengujian akurasi algoritma ML menggunakan dataset SMOTE yang ditampilkan dalam Tabel 4.

Tabel 4. Perbandingan akurasi algoritma ML

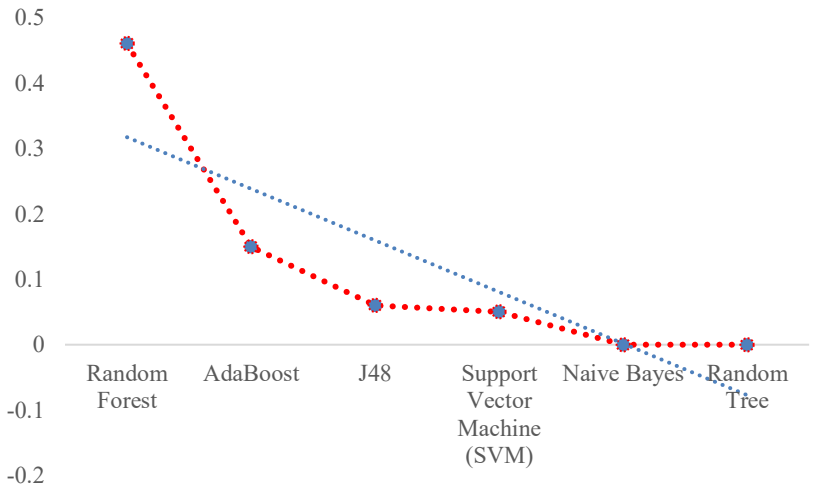
Algoritma	**Awal**	**SMOTE**
Naive Bayes	76.3	73.84
Support Vector Machine (SVM)	77.34	75.28
AdaBoost	74.34	75.38
J48	73.82	75.86
Random Forest	75.78	79.82
Random Tree	68.09	75.19
Rerata	**74.28**	**75.90**

Hasil temuan kami bahwa kombinasi algoritma Random Forest memiliki akurasi terbaik. Hal ini sejalan dengan penelitian [13] bahwa Random Forest memiliki keunggulan dalam akurasi yang tinggi, generalisasi yang baik, dan kemampuan menangani data besar, serta ketahanan terhadap *overfitting*. Selain itu, Random Forest efektif dalam mengelola data yang hilang dan memberikan skor penting fitur, yang berguna untuk pemilihan fitur [15]. Secara umum, penerapan metode keseimbangan *class* mampu meningkatkan akurasi. Penerapan SMOTE meningkatkan akurasi 1.62%. Selanjutnya kami mengukur waktu yang dibutuhkan masing-masing algoritma ML yang ditampilkan pada Gambar 3-4.



Gambar 3. Perbandingan waktu komputasi menggunakan dataset awal

Berdasarkan Gambar 3, temuan kami bahwa Algoritma Naïve Bayes memiliki waktu komputasi terbaik. Namun Gambar 4 menampilkan hasil berbeda ketika dataset sudah dilakukan *balancing class* menggunakan SMOTE. Perbandingan waktu ditampilkan pada Gambar 4.



Gambar 4. Perbandingan waktu komputasi algoritma ML+SMOTE

Gambar 4 menampilkan algoritma dengan waktu komputasi terbaik yaitu Random Tree dan Naïve Bayes. Kedua memiliki memiliki waktu komputasi saat klasifikasi menggunakan dataset SMOTE.

4 KESIMPULAN

Kami melakukan evaluasi terhadap kinerja algoritma ML menggunakan dataset medis PIDD. Dataset ini memiliki jumlah class yang tidak seimbang, sehingga memengaruhi kinerja algoritma ML. Untuk mengatasi hal tersebut, kami

menerapkan metode *oversampling* teknik SMOTE. Penerapan SMOTE mampu meningkatkan jumlah record pada masing-masing class. Selanjutnya kami melakukan klasifikasi dan evaluasi terhadap kinerja algoritma ML

Berdasarkan hasil evaluasi, dapat disimpulkan bahwa ketidakseimbangan kelas dalam dataset diabetes berdampak signifikan terhadap performa model klasifikasi, khususnya dalam hal kemampuan mendeteksi kasus diabetes. Teknik penyeimbangan class menggunakan metode SMOTE terbukti efektif meningkatkan nilai akurasi. Penerapan SMOTE meningkatkan akurasi 1.62% dibandingkan dataset "non-SMOTE" atau dataset awal. Berdasarkan hasil perbandingan, kami menemukan bahwa Algoritma Random Forest memberikan hasil terbaik secara konsisten setelah penerapan teknik *oversampling SMOTE*. Strategi penanganan *imbalanced class* harus menjadi bagian integral dalam *pipeline* pengembangan sistem klasifikasi diabetes, karena dapat mempengaruhi interpretabilitas dan akurasi diagnosis secara keseluruhan.

Hasil perbandingan waktu komputasi algoritma ML menggunakan dataset awal "non-SMOTE" dan SMOTE menunjukkan hasil yang berbeda. Algoritma Random Tree menjadi yang terbaik, diikuti Naive Bayes, sedangkan Random Forest memiliki waktu terburuk. Penelitian lebih lanjut disarankan untuk mengevaluasi efektivitas teknik imbalanced handling lainnya seperti *ensemble-based sampling* atau *cost-sensitive learning*, serta mengaplikasikan pendekatan ini pada dataset klinis yang lebih besar dan komplek

Daftar Pustaka

[1] F. A. Ibrahim and O. A. Shiba, "Data Mining : WEKA Software (an Overview)," *J. Pure Appl. Sci.*, vol. 18, no. 3, pp. 54–58, 2019, [Online]. Available: www.Suj.sebhau.edu.ly

[2] R. B. Lukmanto, Suharjito, A. Nugroho, and H. Akbar, "Early detection of diabetes mellitus using feature selection and fuzzy support vector machine," *Procedia Comput. Sci.*, vol. 157, pp. 46–54, 2019, doi: 10.1016/j.procs.2019.08.140.

[3] W. Thungrut and N. Wattanapongsakorn, "Diabetes classification with fuzzy genetic algorithm," *Adv. Intell. Syst. Comput.*, vol. 769, pp. 107–114, 2019, doi: 10.1007/978-3-319-93692-5_11.

[4] D. Setiawan, A. Nugraha, and A. Luthfiarta, "Komparasi Teknik Feature Selection Dalam Klasifikasi Serangan IoT Menggunakan Algoritma Decision Tree," *J. Media Inform. Budidarma*, vol. 8, no. 1, p. 83, 2024, doi: 10.30865/mib.v8i1.6987.

[5] C. Karima and W. Anggraeni, "Performance Analysis of the Ada-Boost Algorithm For Classification of Hypertension Risk With Clinical Imbalancedd Dataset," *Procedia Comput. Sci.*, vol. 234, pp. 645–653, 2024, doi: https://doi.org/10.1016/j.procs.2024.03.050.

[6] M. F. Ijaz, G. Alfian, M. Syafrudin, and J. Rhee, "Hybrid Prediction Model for type 2 diabetes and hypertension using DBSCAN-based outlier detection, Synthetic Minority Over Sampling Technique (SMOTE), and random forest," *Appl. Sci.*, vol. 8, no. 8, 2018, doi: 10.3390/app8081325.

[7] S. K. Mohapatra, J. K. Swain, and M. N. Mohanty, *Detection of Diabetes Using Multilayer Perceptron*, vol. 846, no. May. Springer Singapore, 2019. doi: 10.1007/978-981-13-2182-5_11.

[8] M. Ohsaki, P. Wang, K. Matsuda, S. Katagiri, H. Watanabe, and A. Ralescu, "Confusion-matrix-based kernel logistic regression for imbalancedd data classification," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1806–1819, 2017, doi: 10.1109/TKDE.2017.2682249.

[9] G. Kovács, "An empirical comparison and evaluation of minority oversampling techniques on a large number of imbalancedd datasets," *Appl. Soft Comput.*, vol. 83, p. 105662, 2019, doi: https://doi.org/10.1016/j.asoc.2019.105662.

[10] A. Myrzakerimova and K. Kolesnikova, "Development of mathematical methods for diagnosing kidney diseases using fuzzy set tools," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 35, no. 1, pp. 405–417, Jul. 2024, doi: 10.11591/ijeecs.v35.i1.pp405-417.

[11] M. Ambika, G. Raghuraman, and L. SaiRamesh, "Enhanced decision support system to predict and prevent hypertension using computational intelligence techniques," *Soft Comput.*, vol. 24, no. 17, pp. 13293–13304, 2020,

doi: 10.1007/s00500-020-04743-9.

[12] J. Nalić, G. Martinović, and D. Žagar, “New hybrid data mining model for credit scoring based on feature selection algorithm and ensemble classifiers,” *Adv. Eng. Informatics*, vol. 45, p. 101130, 2020, doi: https://doi.org/10.1016/j.aei.2020.101130.

[13] H. Sulistiani, A. Syarif, K. Muludi, and Warsito, “Performance evaluation of feature selections on some ML approaches for diagnosing the narcissistic personality disorder,” *Bull. Electr. Eng. Informatics*, vol. 13, no. 2, pp. 1383–1391, 2024, doi: 10.11591/eei.v13i2.6717.

[14] T. Yan, S.-L. Shen, A. Zhou, and X. Chen, “Prediction of geological characteristics from shield operational parameters by integrating grid search and K-fold cross validation into stacking classification algorithm,” *J. Rock Mech. Geotech. Eng.*, vol. 14, no. 4, pp. 1292–1303, 2022, doi: https://doi.org/10.1016/j.jrmge.2022.03.002.

[15] M. Y. Shams, Z. Tarek, and A. M. Elshewey, “A novel RFE-GRU model for diabetes classification using PIMA Indian dataset,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–22, 2025, doi: 10.1038/s41598-024-82420-9.