

Penerapan Teknik SMOTE dalam Mengatasi Imbalance Data Penyakit Diabetes Menggunakan Algoritma ANN

Chanavaro Bayu Handoko¹, Christian Sri Kusuma Aditya²

Program Studi Informatika, Fakultas Teknik, Universitas Muhammadiyah Malang

Email: ¹varohandoko97@webmail.umm.ac.id, ²christianskaditya@umm.ac.id

(Naskah masuk: 30 Juni 2024, diterima untuk diterbitkan: 10 Januari 2025)

Abstrak: Fenomena pada pengujian data machine learning yang sering terjadi adalah data yang tidak seimbang atau sering disebut imbalanced data. Beberapa penelitian lain juga menyampaikan bahwa imbalanced data sering memberikan hasil yang tidak sesuai. Seperti kelas data seringkali mendapatkan nilai ketidakseimbangan yang disebabkan perbedaan nilai rasio antara kelas mayoritas dan minoritas. Cara mengatasi hal tersebut adalah menggunakan teknik oversampling (SMOTE), yang diterapkan pada data diabetes menggunakan algoritma ANN. Validasi keefektifan dari model yang diterapkan, dilakukan dua skema pengujian. Skema pertama, algoritma ANN tanpa oversampling SMOTE yang langsung diterapkan. Pengujian kedua, menggunakan oversampling SMOTE untuk menambah total kelas dataset agar bernilai seimbang. Pengujian yang dilakukan, menghasilkan nilai accuracy, recall, precision terbaik dengan nilai accuracy sebesar 0.92, precision 0.97, dan recall mencapai 0.94 yang menandakan, terbukti efektif dalam memberikan nilai performa pada algoritma Artificial Neural Network jika dibandingkan dengan tidak menerapkan teknik oversampling pada kasus kelas imbalance data diabetes.

Kata Kunci – ANN; Diabetes; Imbalanced Data; SMOTE

Application of SMOTE Technique in Addressing Imbalanced Diabetes Disease Data Using the ANN Algorithm

Abstract: The phenomenon in machine learning data testing that often occurs is unbalanced data or often called imbalanced data. Some other studies also said that imbalanced data often gives inappropriate results. Such data classes often get imbalanced values due to differences in ratio values between majority and minority classes. The way to overcome this is to use the oversampling technique (SMOTE), which is applied to diabetes data using the ANN algorithm. To validate the effectiveness of the applied model, two testing schemes were conducted. The first scheme, ANN algorithm without SMOTE oversampling is directly applied. The second test, using SMOTE oversampling to increase the total class dataset to a balanced value. The tests carried out, resulted in the best accuracy, recall, precision values with an accuracy value of 0.92, precision 0.97, and recall reaching 0.94 which indicates, it is proven effective in providing performance values to the Artificial Neural Network algorithm when compared to not applying oversampling techniques in the case of class imbalance in diabetes data.

Keywords – ANN; Diabetes; Imbalanced Data; SMOTE

1. PENDAHULUAN

Penyakit diabetes merupakan penyakit yang terjadi ketika kondisi metabolisme kronis, ditandai dengan kadar gula yang sangat tinggi di dalam tubuh manusia [1], [2]. Menurut *International Diabetes Federation (IDF)*, pada tahun 2021 lalu tercatat jumlah penderita diabetes telah mencapai 537 juta orang diseluruh dunia. Para peneliti juga memperkirakan bahwa jumlah tersebut akan mengalami peningkatan penderita diabetes menjadi 643 juta pada tahun 2030 [3], [4]. Seseorang yang menderita diabetes, akan mengalami kondisi kadar gula yang berlebih ketika organ pankreas gagal menghasilkan insulin dengan jumlah yang cukup. Diabetes yang tidak ditangani secara medis, dapat mengakibatkan rusaknya berbagai organ tubuh yang lain hingga berdampak kematian [5], [6]. Dengan perkiraan penambahan jumlah penderita tersebut, dapat berpotensi memiliki imbalance data dalam penelitian kasus klasifikasi maupun prediksi.

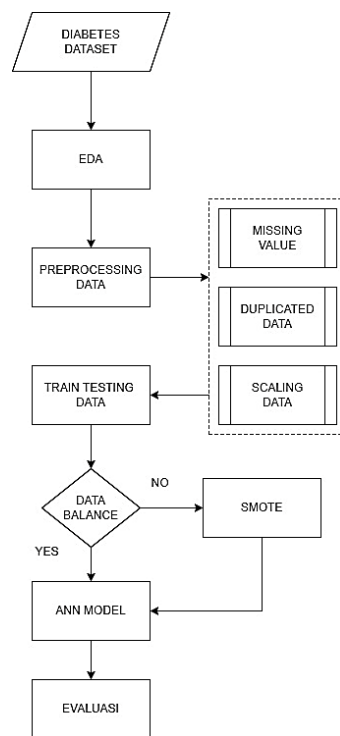
Masalah ketidakseimbangan data sering ditemukan pada berbagai domain, ketika jumlah dalam atribut data tidak sama. Hal ini disebabkan apabila total data dalam 1 atribut secara signifikan lebih banyak atau sedikit dibandingkan atribut data lainnya, sehingga akan mempersulit untuk melakukan proses pengolahan data dan analisis data. Hal tersebut umumnya dikarenakan model analisis data, tidak didesain untuk memperhitungkan distribusi fitur guna peningkatan performa *accuracy* dari model yang digunakan. Dalam mengatasi data yang tidak seimbang, dapat dilakukan dengan menggunakan teknik *resampling*, yakni dengan cara mengubah data pada tiap fitur hingga mencapai jumlah data yang seimbang pada seluruh fitur sebelum melakukan pengujian data [7], [8].

Adapun penerapan teknik *oversampling* dalam beberapa kasus penelitian yang serupa, juga dilakukan seperti implementasi *IQR-SMOTE* dalam klasifikasi diabetes menggunakan algoritma *KNN* (*K-Nearest Neighbors*) dengan hasil performa akurasi 75,78%, recall sebesar 73,88% dan presisi sebesar 63,06% [9]. Penerapan *K-means-SMOTE* dalam klasifikasi diabetes dengan komparasi algoritma *C4.5*, *SVM*, dan *Naive Bayes* memiliki akurasi 0,77% pada algoritma *C4.5*, 0,82% pada *SVM* dan 0,79% algoritma *Naive Bayes* [10]. Selain itu, teknik *Borderline-SMOTE* juga diterapkan pada klasifikasi diabetes dengan algoritma *Bagging-SVM* [11].

Berdasarkan literatur dari penelitian terdahulu, maka penelitian ini akan membangun langkah evaluasi penggunaan teknik *oversampling SMOTE* pada kasus data penyakit diabetes dengan menggunakan algoritma *deep learning* untuk mengklasifikasi, khususnya algoritma *Artificial Neural Network (ANN)*, kemudian dibandingkan dari pemakaian teknik tersebut dengan pengujian tanpa memakai teknik *oversampling*.

2. METODE PENELITIAN

Metodologi yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1. Penelitian dimulai dengan mengumpulkan data diabetes, kemudian dilanjutkan tahap *preprocessing data* sebelum data akan diuji menggunakan algoritma sampai dengan evaluasi.



Gambar 1. Metodologi Penelitian

2.1. Data

Sumber data yang diperoleh dalam penelitian ini, berasal dari *website kaggle* sebanyak 100 ribu record data. Dengan keseluruhan data yang akan diuji memiliki 8 atribut yakni *gender, age, hypertension, heart_disease, smoking_history, bmi, blood_glucose_level, HbA1c_level* (Hemoglobin Terглиkasi) yaitu zat yang terbentuk dari reaksi antara glukosa dan hemoglobin yang bertahan di sel darah merah [12]. Kemudian kelas diabetes sebagai variabel dependen yang menjadi hasil atau respon dari variabel independen.

2.2. Preprocessing Data

Preprocessing data yaitu teknik dari *data mining* dalam mengolah data mentah menjadi data yang menghasilkan informasi [13]. Teknik *preprocessing* yang digunakan meliputi imputasi *missing value* (data yang tidak lengkap), *duplicated data* (data yang bernilai sama), dan normalisasi *StandarScaler* yang menyesuaikan nilai fitur ke skala tertentu [14], [15], [16]. *Preprocessing* diterapkan dalam meningkatkan kualitas data, sehingga analisa yang didapat lebih akurat.

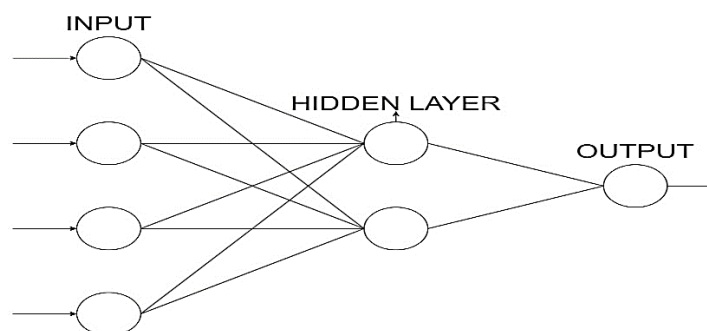
2.3. SMOTE

Synthetic Minority Over-sampling Technique (SMOTE) merupakan teknik *oversampling* untuk menyeimbangkan total data pada kelas minoritas dengan mensintesis atau mereplikasi data minoritas yang baru dari aslinya. Langkah ini dilakukan dengan mengambil *instance* dari kelas minoritas dan menemukan KNN dari masing masing *instance* tersebut, sehingga membantu mengurangi risiko *overfitting* pada model yang digunakan [17], [18].

2.4. ANN

Artificial Neural Network (ANN) merupakan algoritma *deep learning* dengan pola komputasinya berdasarkan sistem jaringan saraf biologis. Algoritma ini tersusun dari beberapa proses sederhana terhubung yang biasa disebut dengan neuron. Proses tersebut dapat terbentuk melalui pemrosesan data, penyesuaian bias sehingga dapat diterapkan dalam masalah prediksi ataupun identifikasi klasifikasi data [19]. Kata '*neural*' mengacu pada unit fungsional dasar atau neuron yang ditemukan pada otak serta komponen lain di tubuh manusia.

Algoritma ANN memiliki sebuah kemampuan dalam mempelajari dan memodelkan dari hubungan *non-linier* dan kompleks. Hal tersebut terdapat peranan utama secara *real time*, karena banyaknya koneksi *input - output* yang tidak linier.



Gambar 2. Struktur Artificial Neural Network

Pada gambar tersebut, *input* akan menerima data yang telah dimasukkan sebelum diteruskan ke dalam *hidden layer*. Setelah itu, *hidden layer* akan melakukan langkah untuk mencocokkan pola dengan cara mengkalkulasi dari data tersebut. Kemudian hasil dari klasifikasi pada *hidden layer* akan dilanjut pada *output layer*[20]. Langkah *training* pada algoritma ANN, akan dijalankan pada saat *i* yang sebagai *neuron* dan *m* sebagai *input*. Pada saat *neuron* berada di *hidden layer* pertama, maka *input* akan terhubung pada *hidden layer* selanjutnya. Pada *neuron output* atau bisa disebut node y_i ,

dapat direpresentasikan dengan y_{ij} dalam mendefinisikan i^{th} sebagai *input* dari *neuron j* di *hidden layer* seterusnya. Adapun rumus aktivasi dari *neuron i* dapat ditulis sebagai berikut [20] :

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

$$y_i = f_i(Net_i) \quad (2)$$

Pada rumus Net_i dijelaskan sebagai keseluruhan pada *node input* dari *neuron i* yang telah dihitung menggunakan persamaan :

$$Net_i = \sum_{j=1}^m w_{i,j} x y_{i,j} + b_{i,0} \quad (3)$$

2.5. Evaluasi Validasi

Pada evaluasi validasi, akan menjelaskan hasil testing dengan data validasi sehingga perbandingan dari nilai actual dengan nilai prediksi dapat dipetakan menggunakan teknik *confusion matrix* yang dapat dirumuskan pada tabel berikut : [21], [22]

Tabel 1. *Confusion Matrix*

		Kelas Prediksi	
		+	-
Kelas / Target	+	TP	FN
	-	FP	TN

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad [23]$$

$$Precision = \frac{TP}{TP+FP} \quad [24]$$

$$Recall = \frac{TP}{TP+FN} \quad [25]$$

3. HASIL DAN PEMBAHASAN

3.1. Pembahasan

Dari tahap ini akan menjelaskan mengenai output pengujian yang telah dilakukan untuk melihat efek penggunaan teknik *SMOTE* terhadap nilai performa dari model ANN. Data diabetes yang telah di lakukan *EDA (Exploratory Data Analysis)*, akan dilakukan *imputasi missing value* sebagai langkah dalam menghindari bias ketika menganalisa dan meningkatkan performa akurasi. Langkah selanjutnya adalah melakukan *duplicated data* yang dimana menghapus entri data sebanyak 17800 yang mempunyai nilai kemiripan dalam dataset, sehingga data akan mudah untuk dikelola.

Untuk memverifikasi keefektifan model yang diimplementasikan, maka dilakukan dua skema pengujian. Pada skema pertama, algoritma ANN tanpa *oversampling SMOTE* diterapkan secara langsung tanpa memperhitungkan keseimbangan kelas. Selanjutnya skema pengujian kedua, menggunakan *oversampling SMOTE* dengan maksud menambah total kelas dataset agar bernilai setara. Perhitungan jumlah data sebelum dan sesudah dilakukan teknik *oversampling SMOTE* dapat dilihat pada tabel berikut :

Tabel 2. Hasil Perhitungan

	Tidak Diabetes	Diabetes
Sebelum	0 : 73821	8379
Sesudah	0 : 73821	73821

Dari tabel tersebut, membuktikan bahwa teknik *oversampling SMOTE* dapat menyeimbangkan nilai minoritas dari kelas diabetes dengan menggunakan data sintetik yang berasal dari replikasi pada data minoritas. Penggabungan algoritma ini, akan menghitung jumlah selisih antara vektor

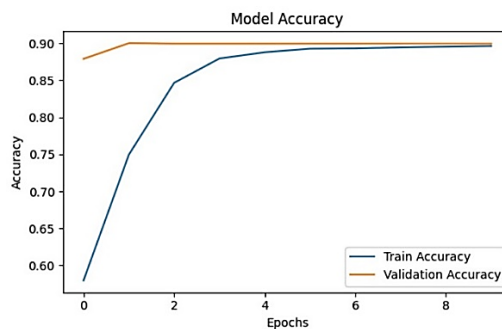
dari fitur kelas minoritas dan jumlah *nearest neighbor*, lalu dikalikan pada bilangan sebelumnya dengan angka acak di antara 0 sampai 1. Hasil akhir dari perhitungan tersebut, selanjutnya digabungkan dengan vektor pada fitur kelas untuk menghasilkan bilangan vektor yang terbaru.

$$X_{new} = X_i + (X_l - X_i) \cdot \delta \quad (4)$$

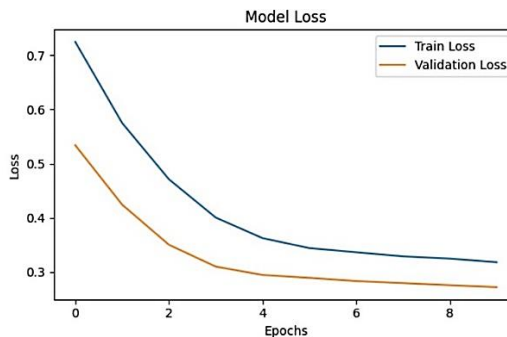
Keterangan :

- X_i = Nilai vektor kelas minoritas
- X_l = Jumlah k - *nearest neighbor* untuk X_i
- δ = Angka acak dari 0 sampai 1

Hasil pada skema pengujian pertama yang disajikan pada gambar 3, menunjukkan model ANN yang diterapkan pada data diabetes cukup optimal, dengan menghasilkan nilai *accuracy* sebanyak 89,63% dan nilai *loss* sebanyak 31,82% pada pengujian *training*. Sementara pada pengujian *validation*, didapatkan nilai untuk *accuracy* sebanyak 89,95% dan *loss* sebanyak 27,22%.

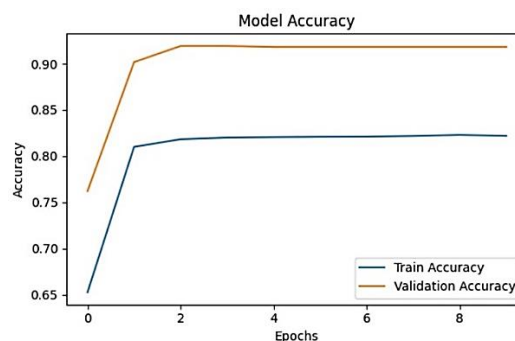


Gambar 3. Grafik *Accuracy* tanpa *smote*

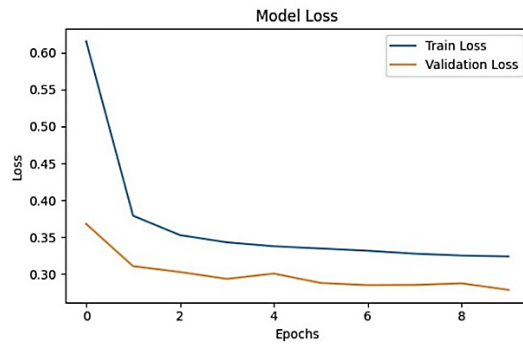


Gambar 4. Grafik *Loss* tanpa *smote*

Kemudian pada pengujian kedua pada gambar 5, menunjukkan model ANN dengan teknik *oversampling SMOTE* dapat menghasilkan nilai peningkatan cukup signifikan dengan nilai *accuracy* yang didapat pada *validation* sebesar 91,82 dan *loss* sebesar 27,88%.

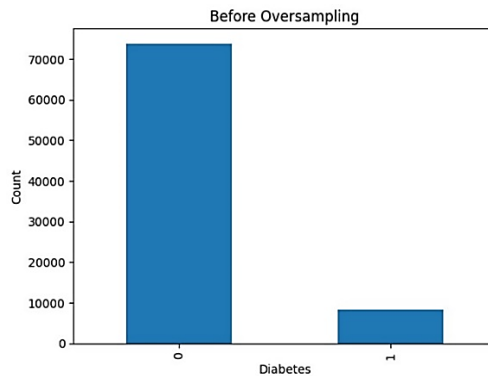


Gambar 5. Grafik *Accuracy* dengan *smote*

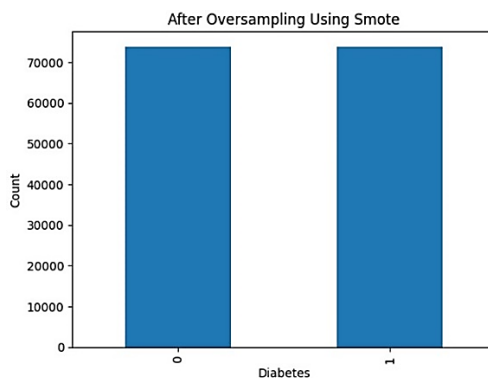


Gambar 6. Grafik Loss dengan smote

Berikut adalah hasil *visual barplot* sebelum dan sesudah penggunaan Teknik *Oversampling SMOTE* yang bisa dilihat pada Gambar 7 dan 8.



Gambar 7. Barplot sebelum *Oversampling*



Gambar 8. Barplot sesudah *Oversampling*

Tabel 3. Metrik Evaluasi

Model Algoritma	Metrik Evaluasi		
	Accuracy (%)	Precision (%)	Recall (%)
ANN	89,95 %	90 %	100 %
ANN + SMOTE	91,82 %	97 %	94 %

Secara keseluruhan, hasil pengujian yang ditunjukkan pada tabel 2, menyatakan bahwa penggabungan Teknik *oversampling* ANN+SMOTE dapat memiliki kinerja yang lebih meningkat dari segi nilai performa akurasi, precision, dan recall jika dibandingkan tanpa menggunakan Teknik *oversampling*.

4. KESIMPULAN

Dalam pengujian ini, penggunaan teknik *oversampling* SMOTE telah dilakukan pada kelas diabetes yang tidak seimbang, dengan mereplikasi data pada kelas diabetes dan tidak diabetes

untuk mendapatkan nilai yang sama. Pengujian data yang telah di *oversampling*, diklasifikasi dengan memanfaatkan model *deep learning Artificial Neural Network (ANN)*. Hasil pengujian serta analisis yang dilakukan, teknik *oversampling* menggunakan *SMOTE* dapat menaikkan nilai performa *accuracy* dalam mengklasifikasi kelas diabetes secara signifikan. Pengujian yang telah dilakukan, menghasilkan nilai *accuracy*, *recall*, *precision* terbaik dengan nilai *accuracy* sebesar 0.92, *precision* 0.97, dan *recall* mencapai 0.94 yang menandakan hampir semua kelas diabetes bisa terdeteksi.

Pengaruh dari penggunaan teknik *oversampling* yang diterapkan, terbukti efektif dalam memberikan nilai performa pada algoritma *Artificial Neural Network* jika dibandingkan dengan tidak menerapkan teknik *oversampling* pada kasus kelas *imbalance* data diabetes.

Selain menggunakan algoritma *Artificial Neural Network (ANN)*, ada banyak algoritma lain yang dapat dieksplorasi dalam menguji teknik *oversampling data*. Penelitian selanjutnya diharapkan dapat mempertimbangkan algoritma *deep learning* sebagai perbandingan meskipun besarnya data yang di *oversampling* memiliki ukuran yang besar, sehingga dalam penelitian ini masih dapat terus dikembangkan.

UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada para dosen Program Studi Informatika Universitas Muhammadiyah Malang yang telah mendukung kelancaran penelitian ini, sehingga membantu penulis dalam mengembangkan penelitian ini dan bermanfaat untuk kedepannya.

DAFTAR PUSTAKA

- [1] Dikan Ismafillah, Tatang Rohana, and Yana Cahyana, "Analisis algoritma pohon keputusan untuk memprediksi penyakit diabetes menggunakan *oversampling smote*," *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 27–36, 2023, doi: [10.37373/infotech.v4i1.452](https://doi.org/10.37373/infotech.v4i1.452).
- [2] M. Ardiansyah, A. Sunyoto, and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes Dan C4.5 untuk Klasifikasi Diabetes," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 2, pp. 147–156, 2021, doi: [10.29408/edumatic.v5i2.3424](https://doi.org/10.29408/edumatic.v5i2.3424).
- [3] "International Diabetes Federation." <https://idf.org/about-diabetes/diabetes-facts-figures/> (accessed Mar. 20, 2024).
- [4] W. Nugraha and R. Sabaruddin, "Teknik Resampling untuk Mengatasi Ketidakseimbangan Kelas pada Klasifikasi Penyakit Diabetes Menggunakan C4.5, Random Forest, dan SVM," *Techno.Com*, vol. 20, no. 3, pp. 352–361, 2021, doi: [10.33633/tc.v20i3.4762](https://doi.org/10.33633/tc.v20i3.4762).
- [5] U. E. Erdiansyah, "Klasifikasi Penyakit Diabetic Retinopathy Menggunakan Multilayer Perceptron," *J. Artif. Intell. Softw. Eng.*, vol. 2, no. 1, pp. 1–6, 2022, doi: [10.30811/jaise.v2i1.3084](https://doi.org/10.30811/jaise.v2i1.3084).
- [6] 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.ict.2021.02.004](https://doi.org/10.1016/j.ict.2021.02.004).
- [7] A. Indrawati, "Penerapan Teknik Kombinasi *Oversampling* Dan *Undersampling* Hybrid *Oversampling* and *Undersampling* Techniques To Handling Imbalanced Dataset," *JIKO (Jurnal Inform. dan Komputer)*, vol. 4, no. 1, pp. 38–43, 2021.
- [8] D. Mualfah *et al.*, "Jurnal Computer Science and Information Technology (CoSciTech) algoritma random forest," vol. 3, no. 2, pp. 107–113, 2022.
- [9] S. M. Faradisa, T. D. Nugrahadi, Muliadi, I. Budiman, and D. Kartini, "Implementasi IQR-SMOTE Untuk Mengatasi Ketidakseimbangan Kelas Pada Klasifikasi Diabetes menggunakan K-Nearest Neighbors," vol. 15, pp. 48–60, 2021.
- [10] H. Hairani, K. E. Saputro, and S. Fadli, "K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 2, pp. 89–93, 2020, doi: [10.14710/jtsiskom.8.2.2020.89-93](https://doi.org/10.14710/jtsiskom.8.2.2020.89-93).

- [11] Trianto, A. Muliawati, and H. N. Irmanda, "Penerapan Borderline-SMOTE dan Grid Search pada Bagging-SVM untuk Klasifikasi Penyakit Diabetes," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, vol. 3, no. 2, pp. 102–113, 2022.
- [12] T. Rawung, J. Posangi, and E. Nangoy, "Efektivitas Penggunaan Empagliflozin terhadap Nilai HbA1c pada Pasien Diabetes Melitus Tipe 2," *Med. Scope J.*, vol. 5, no. 2, pp. 232–239, 2023, doi: [10.35790/msj.v5i2.45424](https://doi.org/10.35790/msj.v5i2.45424).
- [13] Aditya Gumilar, Sri Suryani Prasetiyowati, and Yuliant Sibaroni, "Performance Analysis of Hybrid Machine Learning Methods on Imbalanced Data (Rainfall Classification)," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 3, pp. 481–490, 2022, doi: [10.29207/resti.v6i3.4142](https://doi.org/10.29207/resti.v6i3.4142).
- [14] M. R. A. Prasetya, A. M. Priyatno, and Nurhaeni, "Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, vol. 5, no. 2, pp. 52–62, 2023, doi: [10.37034/jidt.v5i2.324](https://doi.org/10.37034/jidt.v5i2.324).
- [15] M. R. Andryan, M. Fajri, and N. Sulistyowati, "Komparasi Kinerja Algoritma Xgboost Dan Algoritma Support Vector Machine (Svm) Untuk Diagnosis Penyakit Kanker Payudara," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 1, p. 1, 2022, doi: [10.26798/jiko.v6i1.500](https://doi.org/10.26798/jiko.v6i1.500).
- [16] E. F. L. Awalina and W. I. Rahayu, "Optimalisasi Strategi Pemasaran dengan Segmentasi Pelanggan Menggunakan Penerapan K-Means Clustering pada Transaksi Online Retail," *J. Teknol. dan Inf.*, vol. 13, no. 2, pp. 122–137, 2023, doi: [10.34010/jati.v13i2.10090](https://doi.org/10.34010/jati.v13i2.10090).
- [17] G. Gumelar, Q. Ain, R. Marsuciati, S. Agustanti Bambang, A. Sunyoto, and M. Syukri Mustafa, "Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance," *SISFOTEK Sist. Inf. dan Teknol.*, pp. 250–255, 2021.
- [18] A. Wibowo and S. Rahmawati, "Evaluasi Model Klasifikasi Algoritma Terbimbing Kuantitatif terhadap Penyakit Diabetes," vol. 7, no. 1, pp. 127–134, 2023.
- [19] E. Sutoyo and M. A. Fadlurrahman, "Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Television Advertisement Performance Rating Menggunakan Artificial Neural Network," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 3, p. 379, 2020, doi: [10.26418/jp.v6i3.42896](https://doi.org/10.26418/jp.v6i3.42896).
- [20] F. Handayani *et al.*, "JEPIN (Jurnal Edukasi dan Penelitian Informatika) Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network dalam Prediksi Penyakit Jantung," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 3, p. Vol. 7 No. 3, 2021.
- [21] L. A. Andika, P. A. N. Azizah, and R. Respatiwan, "Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier," *Indones. J. Appl. Stat.*, vol. 2, no. 1, p. 34, 2019, doi: [10.13057/ijas.v2i1.29998](https://doi.org/10.13057/ijas.v2i1.29998).
- [22] Nurul Chamidah, Mayanda Mega Santoni, and Nurhafifah Matondang, "Oversampling Method on Classifying Hypertension Using Naive Bayes, Decision Tree, and Artificial Neural Network," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 4, pp. 635–641, 2020, doi: [10.29207/resti.v4i4.2015](https://doi.org/10.29207/resti.v4i4.2015).
- [23] D. Pramadhana, "Klasifikasi Penyakit Diabetes Menggunakan Metode CFS dan ROS dengan Algoritma J48 Berbasis Adaboost," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 89–98, 2021, doi: [10.29408/edumatic.v5i1.3336](https://doi.org/10.29408/edumatic.v5i1.3336).
- [24] W. Nugraha and M. Syarif, "Teknik Weighting untuk Mengatasi Ketidakseimbangan Kelas Pada Prediksi Churn Menggunakan XGBoost, LightGBM, dan CatBoost," *Techno.Com*, vol. 22, no. 1, pp. 97–108, 2023, doi: [10.33633/tc.v22i1.7191](https://doi.org/10.33633/tc.v22i1.7191).
- [25] I. Y. R. Pratiwi, "Hoax news identification using machine learning model from online media in Bahasa Indonesia," *MATRIX J. Manaj. Teknol. dan Inform.*, vol. 12, no. 2, pp. 58–67, 2022, doi: [10.31940/matrix.v12i2.58-67](https://doi.org/10.31940/matrix.v12i2.58-67).