

Terakreditasi SINTA Peringkat 3

Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, dan Teknologi Nomor 225/E/KPT/2022 masa berlaku mulai Vol.7 No. 1 tahun 2022 s.d Vol. 11 No. 2 tahun 2026

Terbit online pada laman web jurnal:
<http://publishing-widyagama.ac.id/ejournal-v2/index.php/jointecs>



Vol. 8 No. 3 (2023) 127 - 134

JOINTECS

(Journal of Information Technology and Computer Science)

e-ISSN:2541-6448

p-ISSN:2541-3619

Evaluasi Model Klasifikasi Algoritma Terbimbing Kuantitatif terhadap Penyakit Diabetes

Arief Wibowo¹, Selly Rahmawati²

Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan

¹arief.wibowo@budiluhur.ac.id, ²2211601055@student.budiluhur.ac.id

Abstract

Diabetes is a medical condition characterized by a persistent metabolic disorder resulting in elevated glucose levels in the bloodstream. This ailment has a profound impact on various bodily organs, including the heart, blood vessels, eyes, kidneys, and nervous system. One significant factor contributing to the rise in diabetes cases is the delay in diagnosing the condition. The objective of this study was to assess different algorithms for the detection of diabetes. The research involved an imbalanced dataset, necessitating the application of oversampling techniques like Synthetic Minority Over-sampling Technique (SMOTE) to address this imbalance. Two classification methods, namely Support Vector Machine (SVM) and Logistic Regression (LR), were employed in this investigation. The study findings revealed that when the K-Fold Cross Validation technique was combined with the SMOTE method, the Support Vector Machine (SVM) model exhibited superior levels of accuracy, precision, and recall in comparison to the Logistic Regression (LR) model, also utilizing the SMOTE technique. Nonetheless, if K-Fold Cross Validation was conducted without implementing the SMOTE technique, the results demonstrated that Logistic Regression outperformed the Support Vector Machine (SVM) model in terms of accuracy, precision, and recall.

Keywords: Diabetes; Support Vector Machine; Logistic Regression; K-Fold Cross Validation.

Abstrak

Diabetes adalah suatu kondisi medis yang ditandai dengan gangguan metabolisme persisten yang mengakibatkan peningkatan kadar glukosa dalam aliran darah. Penyakit ini berdampak besar pada berbagai organ tubuh, termasuk jantung, pembuluh darah, mata, ginjal, dan sistem saraf. Salah satu faktor penting yang berkontribusi terhadap peningkatan kasus diabetes adalah keterlambatan diagnosis kondisi tersebut. Tujuan dari penelitian ini adalah untuk menilai algoritma yang berbeda untuk mendeteksi diabetes. Penelitian ini melibatkan kumpulan data yang tidak seimbang, sehingga memerlukan penerapan teknik oversampling seperti Synthetic Minority Over-sampling Technique (SMOTE) untuk mengatasi ketidakseimbangan ini. Dua metode klasifikasi, yaitu Support Vector Machine (SVM) dan Logistic Regression (LR), digunakan dalam penyelidikan ini. Temuan penelitian mengungkapkan bahwa ketika teknik K-Fold Cross Validation dikombinasikan dengan metode SMOTE, model Support Vector Machine (SVM) menunjukkan tingkat akurasi, presisi, dan recall yang lebih unggul dibandingkan dengan model Logistic Regression (LR), juga menggunakan teknik SMOTE. Meskipun demikian, jika K-Fold Cross Validation dilakukan tanpa menerapkan teknik SMOTE, hasilnya menunjukkan bahwa Regresi Logistik mengungguli model Support Vector Machine (SVM) dalam hal akurasi, presisi, dan recall.

Kata kunci: Penyakit diabetes; Support Vector Machine; Logistic Regression; K-Fold Cross Validation.

Diterima Redaksi : 23-08-2023 | Selesai Revisi : 16-09-2023 | Diterbitkan Online : 30-09-2023



1. Pendahuluan

Menurut Federasi Diabetes Internasional (2019), Indonesia berada di peringkat ke-7 dalam hal prevalensi penyakit diabetes, dengan 10,7% dari populasi Indonesia [1]. Selain itu, menurut WHO sekitar 422 juta orang di seluruh dunia menderita diabetes kebanyakan penderita diabetes tinggal di negara-negara berpenghasilan rendah dan menengah, dan hingga sebanyak 1,5 juta kematian secara langsung dikaitkan dengan diabetes setiap tahun [2].

Menurut data yang dirilis oleh International Diabetes Federation (IDF), pada tahun 2022, Indonesia memiliki sekitar 41,8 ribu individu yang menderita diabetes tipe 1. Hal ini menjadikan Indonesia sebagai negara di ASEAN dengan jumlah penderita diabetes tipe 1 tertinggi, dan juga menempatkannya pada peringkat ke-34 dari 204 negara di seluruh dunia. Dampak dari kondisi diabetes dapat mengakibatkan kerusakan pada organ-organ tubuh seperti jantung, pembuluh darah, mata, ginjal, dan saraf. [3]. Adapun jumlah kasus maupun prevalensi diabetes ini terus meningkat selama beberapa dekade terakhir [4]. Sehingga alokasi anggaran Kesehatan dunia berkembang hanya untuk menangani diabetes.

Diabetes merupakan suatu penyakit kronis dan *metabolic* yang ditandai oleh kadar glukosa darah atau gula darah yang tinggi [5]. Penyebabnya adalah gaya hidup yang tidak sehat dan pola makan yang tidak seimbang [6]. Usia terbanyak penderita penyakit diabetes adalah antara 49-59 tahun [7]. Salah satu faktor yang menyebabkan peningkatan jumlah penderita diabetes adalah keterlambatan dalam mendiagnosis penyakit diabetes [8]. Oleh karena itu, diperlukan Teknik data mining untuk membantu para medis dalam melakukan diagnosa dini atau klasifikasi berdasarkan variabel-variabel riwayat pasien diabetes sebelumnya. Adapun Teknik data mining ini terbukti mampu memberikan dampak secara ilmiah pada penelitian sebelumnya [9], [5], [10], [11].

Studi terdahulu, telah melakukan pengklasifikasian penyakit diabetes menggunakan *machine learning* oleh Ichwanul Muslim et.al [9]. Pada penelitiannya menggunakan SVM, Decision Tree dan Naïve bayes. Hasil penelitian ini mengindikasikan bahwa SVM adalah algoritma dengan kinerja terbaik. Selain itu, Agung Mulyo dkk. melakukan perbandingan performa algoritma K-NN, J48, Naïve Bayes, dan Regresi Logistik dalam klasifikasi penyakit diabetes dalam penelitian mereka sebelumnya [12]. Pada penelitiannya menunjukkan algoritma K-NN menghasilkan *accuracy* tertinggi yaitu 98%. Selanjutnya, penelitian yang telah dilakukan oleh Ginanjar Abdurrahman dengan menggunakan *Adaboost Classifier* untuk

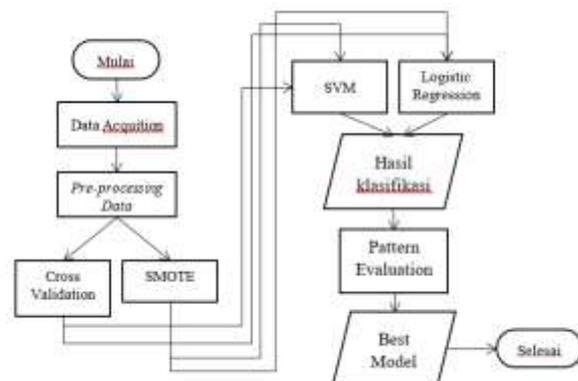
mengklasifikasikan penyakit diabetes [10]. Pada penelitiannya menggunakan Teknik *imputing mean* dan *median* untuk proses pengujiannya. Hasil penelitiannya menunjukkan bahwa *accuracy* tertinggi didapatkan dengan menggunakan Teknik *imputing mean* yaitu 80,09%. Penelitian selanjutnya, Nunu et.al melakukan komparasi algoritma ID3 dan Naïve Bayes untuk klasifikasi Diabetes.

Pada penelitiannya menggunakan pemrograman python diaplikasi Jupiter Notebook. Berdasarkan penelitian yang telah dilakukan, algoritma Naïve bayes memiliki hasil prediksi yang lebih tinggi dibandingkan ID3 [11]. Kemudian, hilda et.al juga melakukan perbandingan metode Naïve bayes dan Support Vector Machine dalam klasifikasi Diabetes. Pada penelitiannya menggunakan lara bantu WEKA. Adapun hasil dari penelitian tersebut menunjukkan bahwa algoritma SVM dengan kernel polynomial lebih baik dari pada Naïve Bayes [13]. Selanjutnya, Naisah et.al juga melakukan komparasi algoritma KNN DAN Naïve bayes pada klasifikasi penyakit Diabetes. Studi lain yang menggunakan tahapan KDD memiliki nilai akurasi paling tinggi pada algoritma Naïve bayes dibanding dengan KNN [14].

Berdasarkan sumber dari beberapa penelitian menjelaskan bahwa model klasifikasi dapat membantu layanan kesehatan dalam melakukan penanganan dan pencegahan terhadap penderita penyakit diabetes. Di sisi lain, model klasifikasi juga mampu melakukan prediksi yang akurat. Oleh karena itu, penelitian ini bertujuan untuk meningkatkan efektivitas layanan kesehatan dalam menangani penyakit diabetes dengan melakukan perbandingan kinerja antara algoritma *Support Vector Machine (SVM)* dan *Logistic Regression (LR)* dalam klasifikasi penyakit diabetes.

2. Metode Penelitian

Pada tahap ini, dijelaskan mengenai proses dan tahapan metode penelitian yang digunakan. Tahapan penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur Penelitian

Penelitian ini dimulai dengan langkah Akuisisi Data yang berasal dari Kaggle sebagai sumber data. Setelah itu, proses berlanjut ke tahap pengolahan data, di mana data diolah menggunakan teknik *pre-processing*. Selanjutnya, data diklasifikasikan metode untuk mengklasifikasikan penyakit diabetes, yaitu *Support Machine Learning* (SVM) dan *Logistic Regression* (LR). Selanjutnya, dilakukan evaluasi terhadap dua metode tersebut untuk menentukan kinerja yang paling optimal, sehingga menghasilkan performa terbaik.

2.1 Data Acquisition

Data Acquisition merupakan suatu tahapan untuk pengambilan data. Data yang digunakan pada penelitian ini yaitu data pasien penderita penyakit diabetes sebanyak 2.000 objek dengan 8 variabel dan 1 kelas target. Data yang digunakan adalah data kuantitatif berupa angka-angka yang akan diolah menggunakan metode statistika. Sumber data diambil dari catatan medis yang disetujui oleh Organisasi Kesehatan Dunia dan tersedia pada tahun 2022. Variabel-variabel yang dianalisis meliputi *Pregnancies*, *Glucose*, *BloodPressure*, *SkinThickness*, *Insulin*, *BMI*, *Diabetes Pedigree Function* dan *Age*.

2.2 Preprocessing

Pada tahapan *preprocessing* penelitian ini mengimplementasikan empat fase analitik (analitik kualitas data, analitik deskriptif, analitik diagnostik, dan analitik prediktif). Fase pertama meliputi pra-pemrosesan data termasuk analitik kualitas data dan analitik deskriptif. Setelah itu, hasil dari pra-pemrosesan data diproses untuk menghasilkan analitik diagnostik dan analitik prediktif. Adapun pada tahapan yang digunakan yaitu seleksi fitur, cleaning data dan normalisasi data. Adapun Seleksi fitur merupakan suatu tahapan untuk pemilihan atribut yang akan digunakan pada penelitian. Sedangkan Cleaning data merupakan suatu tahapan untuk menghilangkan data dari noise ataupun missing value [15]. Normalisasi data merupakan suatu proses untuk merubah penskalaan data, sehingga proses pengolahan menjadi lebih mudah [16].

2.3 Klasifikasi Penyakit Diabetes

Dalam penelitian ini, akan menerapkan teknik oversampling SMOTE. SMOTE (Synthetic Minority Over-sampling Technique) merupakan suatu metode oversampling yang bertujuan untuk meningkatkan jumlah data pada kelas minoritas dengan menciptakan data sintetis melalui replikasi data kelas minoritas. Dalam SMOTE, *over-sampling* dilakukan dengan mengambil instance dari kelas minoritas dan menemukan *K-Nearest Neighbor* dari masing-masing instance tersebut. Kemudian, *instance* sintetis dibuat dengan mempertimbangkan instance-instance tersebut, menghindari replikasi langsung dari instance kelas

minoritas. Dengan demikian, metode ini dapat membantu mengatasi masalah *overfitting* yang berlebihan [17].

SVM (Support Vector Machine) merupakan suatu metode pembelajaran mesin yang dirancang untuk mengidentifikasi hyperplane optimal yang dapat memisahkan dua kelas dalam ruang masukan [18]. Konsep dasar SVM merupakan salah satu metode yang termasuk dalam kategori metode terbimbing. Metode terbimbing mengacu pada metode yang memerlukan penggunaan data *training* dan data *testing* dalam proses pengujian atau evaluasi [19]. *Logistic Regression* adalah metode untuk menentukan hubungan antara variabel respons yang memiliki sifat dikotomis atau polikotomis [20].

2.4 Evaluasi

Evaluasi dilakukan dalam penelitian ini berdasarkan *confusion matrix*. *Confusion matrix* merupakan tabel yang menunjukkan klasifikasi jumlah data uji yang benar dan jumlah data uji yang salah [21]. Adapun rumus yang digunakan untuk menghitung *accuracy*, *precision* dan *recall* sebagai berikut:

$$Accuracy = \frac{TP+TN}{Total} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

Selain itu, dalam penelitian ini dilakukan evaluasi dengan menggunakan metode *K-Fold Cross Validation* dalam pengujian. *K-Fold Cross validation* merupakan metode teknik penambangan data yang bertujuan untuk mencapai akurasi maksimum dengan membagi data menjadi dua subset, yaitu data pelatihan dan data pengujian [22]. Dalam klasifikasi, model dilatih dan diuji dalam k iterasi. Pada setiap iterasi, salah satu subset digunakan sebagai data pelatihan dan data pengujian [23].

3. Hasil dan Pembahasan

Pada tahapan penelitian ini, hasil dan pembahasan akan dijelaskan berdasarkan alur penelitian. Alur penelitian dimulai dari data acquisition, preprocessing, modeling, evaluasi serta perbandingan model. Adapun penjelasan mengenai hasil dan pembahasan dari setiap langkah alur penelitian sebagai berikut:

3.1 Pre-processing

Pada tahap preprocessing dilakukan beberapa tahapan. Tahapan pertama yang dilakukan yaitu mengubah tipe data. Perubahan tipe data ini bertujuan agar mudah dipahami dan meningkatkan efisiensi dalam pemrosesan data. Tipe data yang diubah pada tahap ini yaitu int menjadi *category*. Adapun perubahan tipe data dilihat pada Gambar 2.

Age	2000 non-null	int64
Outcome	2000 non-null	int64

Gambar 2. tipe data sebelum diubah

Age	int64
Outcome	category

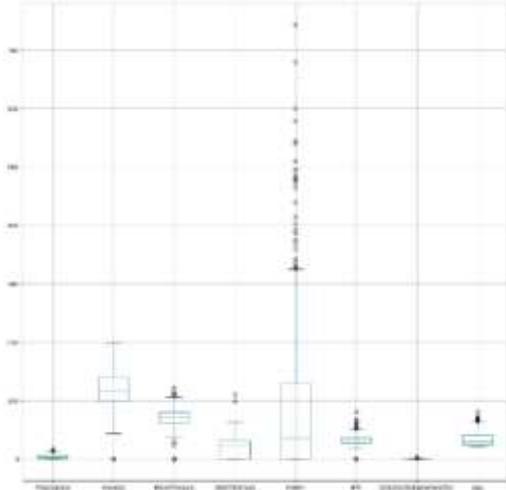
Gambar 3. tipe data setelah diubah

	Pregnancies	Glucose	BloodPressure	SkinfoldThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	2	138	62	35	9	33.6	0.167	47
1	8	94	80	31	125	30.2	0.233	21
3	8	138	60	42	208	42.1	0.360	24
8	8	173	78	32	204	46.4	0.199	68
6	4	94	52	47	8	28.4	0.254	28

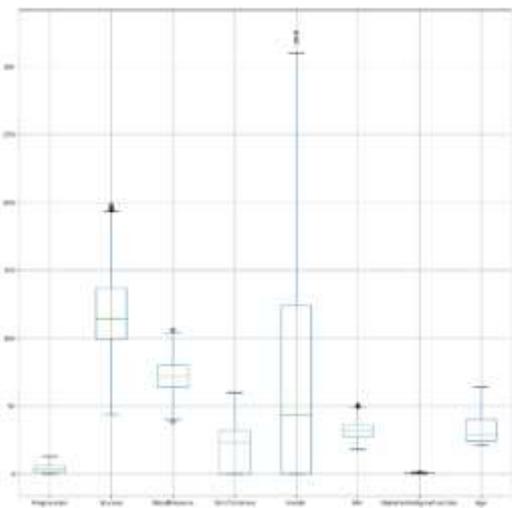
Gambar 6. Sebelum Normalisasi Data

	Pregnancies	Glucose	BloodPressure	SkinfoldThickness	Insulin	BMI	DiabetesPedigreeFunction	Age
0	0.000000	0.000000	0.000000	0.000000	0.000000	0.473848	0.044020	0.040000
1	0.000000	0.207407	0.000000	0.196877	0.402226	0.611264	0.139263	0.040000
3	0.000000	0.500000	0.000000	0.000000	0.000000	0.741978	0.257082	0.000000
8	0.000000	0.237862	0.000000	0.000000	0.000000	0.679768	0.071290	0.000000
6	0.000000	0.000000	0.000000	0.000000	0.000000	0.227000	0.184070	0.040000

Gambar 7. Sesudah Normalisasi Data



Gambar 4. Sebelum Pembersihan Data



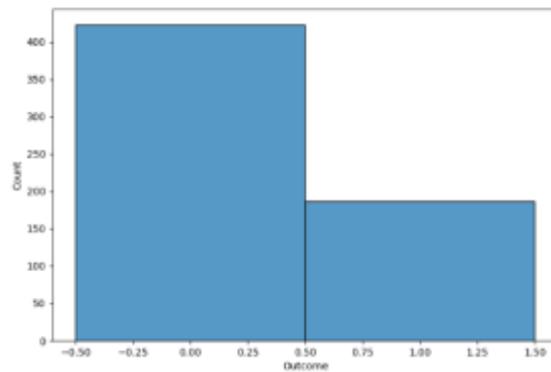
Gambar 5. Setelah Pembersihan Data

Terlihat pada Gambar 3 bahwa setelah dilakukan seleksi data pada penelitian ini terjadi perubahan nilai data. Seleksi data yang dilakukan menggunakan *correlation based fitur*. Kemudian, dilakukan *cleaning* data. *cleaning* data digunakan untuk mengatasi data dari *missing value*, duplikat data, *noise* dan *outlier*. Adapun proses pembersihan (*cleaning*) data dapat dilihat pada Gambar 4 dan Gambar 5.

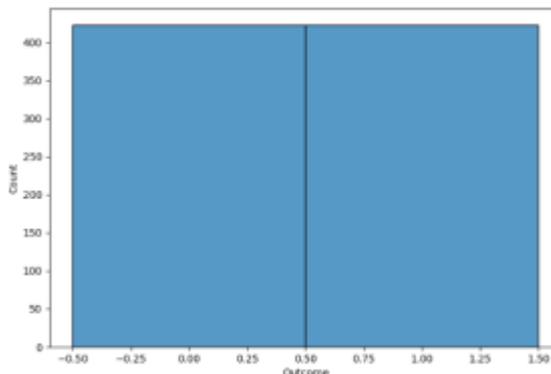
Setelah dilakukan *cleaning* data, dilanjut dengan normalisasi data. Tahapan normalisasi data ini dilakukan guna untuk menghilangkan perbedaan skala antar atribut. Adapun normalisasi data dapat dilihat pada Gambar 6.

3.2 Pemodelan

Data yang diperoleh dari sumber *Kaggle* tidak seimbang. Oleh karena itu, penelitian ini perlu diterapkannya suatu Teknik *oversampling* yaitu SMOTE. Adapun data sebelum dan sesudah menggunakan teknik SMOTE dapat dilihat pada gambar dibawah ini:



Gambar 8. Before SMOTE



Gambar 9. After SMOTE

Setelah menerapkan Teknik SMOTE, langkah berikutnya adalah menerapkan model dengan melakukan pembagian dataset menggunakan K-Fold Cross Validation. Setelah dataset dipartisi, Langkah selanjutnya adalah melakukan pemodelan dengan menerapkan algoritma *Support Vector Machine (SVM)* dan *Logistic Regression (LR)*. Setelah itu, evaluasi dilakukan dengan membandingkan kinerja SVM dan *Logistic Regression (LR)* pada dataset yang tidak mengalami pemrosesan dengan teknik SMOTE dan dataset yang telah melalui proses teknik SMOTE.

Tabel 1. Hasil *K-Fold* 1 Sampai 10 dengan SMOTE

K-Fold	SVM			Logistic Regression		
	Akurasi	Precision	Recall	Akurasi	Precision	Recall
1	88%	81%	94%	80%	71%	92%
2	80%	87%	78%	73%	87%	65%
3	79%	75%	79%	78%	76%	74%
4	82%	81%	79%	75%	69%	82%
5	79%	85%	74%	78%	86%	69%
6	82%	85%	79%	75%	82%	65%
7	76%	82%	71%	76%	80%	73%
8	71%	76%	73%	68%	76%	65%
9	82%	77%	87%	76%	71%	79%
10	83%	81%	81%	78%	73%	81%
Rata-Rata	80%	81%	80%	75%	77%	74%

Tabel 2. Hasil *K-Fold* 1 Sampai 10 Tanpa SMOTE

K-Fold	SVM			Logistic Regression		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
1	77%	86%	31%	72%	62%	26%
2	72%	82%	38%	72%	89%	33%
3	75%	70%	37%	75%	67%	42%
4	85%	85%	63%	85%	86%	63%
5	72%	65%	50%	77%	75%	54%
6	84%	86%	60%	79%	77%	50%
7	79%	57%	53%	82%	64%	60%
8	82%	73%	61%	82%	73%	61%
9	85%	73%	57%	80%	58%	50%
10	78%	71%	31%	80%	75%	37%
Rata-Rata	79%	75%	48%	78%	72%	48%

Tabel 3. Perbandingan Performa

Metode	Accuracy	Precision	Recall
SVM	78%	71%	31%
LR	80%	75%	37%
SVM + SMOTE	83%	81%	81%
LR + SMOTE	78%	73%	81%

3.3 Evaluasi

Penggunaan metode validasi silang K-Fold dalam pengujian dilakukan untuk mengurangi distorsi yang mungkin terjadi dengan data acak [24]. Dalam penelitian ini, K-Fold 10 kali lipat dilakukan dengan membagi kumpulan data menjadi 10 lipatan dengan ukuran yang sama. Selanjutnya, dilakukan pelatihan dan pengujian pada model sebanyak 10 kali. Kemudian, klasifikasi dilakukan menggunakan algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) dengan memanfaatkan metode SMOTE. Berikut hasil K-Fold 1 sampai 10 yang didapatkan pada Tabel 1,

Berdasarkan Tabel 1, diperoleh bahwa *accuracy* dan *recall* tertinggi terdapat pada K-Fold 1 dengan *accuracy* SVM 88%, *Recall* 94% dan LR 80%, *Recall* 92%. Sedangkan *Precision* tertinggi terdapat pada K-Fold 2 dengan SVM 97% dan LR 87%. Tahap selanjutnya penerapan algoritma SVM dan *Logistic Regression* tanpa menggunakan teknik SMOTE pada K-Fold 1 sampai K-Fold 10 dengan hasil pada Tabel 2. Berdasarkan Tabel 2, diperoleh bahwa *accuracy* dan

recall tertinggi terdapat pada K-Fold 4 dengan *accuracy* SVM 85%, *recall* 85% dan LR 85%, *recall* 86%. Sedangkan *Precision* tertinggi terdapat pada K-Fold 6 dengan SVM 86% dan *Precision* dengan LR tertinggi terdapat pada K-Fold 2 yaitu 89%.

3.4 Perbandingan Model

Berdasarkan pada penelitian sebelumnya [25] menunjukkan diketahui bahwa K-Fold 10 menjadi pemodelan terbaik. Ujicoba pada studi ini juga menggunakan K-Fold 10. Adapun hasil evaluasi tersebut ditunjukkan pada Tabel 3. Berdasarkan hasil penelitian menggunakan K-Fold 10 didapatkan nilai rata-rata *precision* dan *recall* yang lebih rendah dibandingkan dengan akurasi. Meskipun nilai rata-rata nilai *precision* dan *recall* rendah, namun model tersebut terbukti mendapatkan hasil akurasi yang baik. Kemudian, model klasifikasi dengan menggunakan teknik SMOTE mendapatkan kenaikan dibandingkan dengan model tanpa SMOTE. Pada tabel 3 menunjukkan bahwa SVM dengan menggunakan Teknik SMOTE lebih unggul dibandingkan dengan *Logistic regression*. SVM dengan menggunakan Teknik SMOTE mendapat nilai *accuracy* 83%, *precision* 81% dan *recall* 81%. Sedangkan *Logistic Regression* dengan menggunakan Teknik SMOTE mendapat nilai *accuracy* 78%, *precision* 73% dan *recall* 81%. Adapun perbandingan tanpa Teknik SMOTE, *Logistic Regression* lebih unggul dibandingkan dengan SVM. Adapun *Logistic Regression* mendapat nilai *accuracy* 80%, *precision* 75% dan *recall* 37%. Sedangkan SVM mendapat nilai *accuracy* 78%, *precision* 71% dan *recall* 31%.

4. Kesimpulan

Hasil penelitian menunjukkan bahwa SVM dengan menggunakan Teknik SMOTE memiliki nilai *accuracy*, *precision*, *recall* yang lebih tinggi dibandingkan dengan *Logistic Regression*. Adapun SVM mendapatkan nilai *accuracy* 83%, *precision* 81%, dan *recall* 81%. Sedangkan *Logistic Regression* mendapat nilai *accuracy* 78%, *precision* 73% dan *recall* 81%. Akan tetapi perbandingan tanpa Teknik SMOTE *Logistic Regression* lebih unggul dibandingkan dengan SVM. Adapun *Logistic Regression* mendapat nilai *accuracy* 80%, *precision* 75% dan *recall* 37%. Sedangkan SVM mendapat nilai *accuracy* 78%, *precision* 71% dan *recall* 31%. Pada penelitian yang telah dilakukan ini hanya membandingkan dua algoritma klasifikasi yaitu *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Diharapkan pada penelitian selanjutnya dapat menggunakan algoritma lain, sehingga mendapatkan performa yang lebih baik dengan menggunakan dataset diabetes.

Daftar Pustaka

- [1] C. D. Mait, J. A. Watuseke, P. D. G. Saerang, and S. R. Joshua, "Sistem Pendukung Keputusan

- Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan Penyakit Diabetes,” *J. Media Infotama*, vol. 18, no. 2, p. 344, 2022.
- [2] C. Mela and A. Barkah, “Hubungan Dukungan Keluarga Dengan Kepatuhan Menjalani Diet Pada Pasien Diabetes Melitus Di Di Jorong Koto Kaciak Nagari Batu Balang Kecamatan Harau ...,” *J. Pendidik. dan ...*, vol. 4, no. 1716–1724, pp. 1716–1724, 2022, [Online]. Available: <http://journal.universitaspahlawan.ac.id/index.php/jpdk/article/view/4949%0Ahttp://journal.universitaspahlawan.ac.id/index.php/jpdk/article/download/4949/3403>
- [3] D. Pademme and T. Banna, “Peer Group Support Terhadap Self- Efficacy Pasien DM Tipe II,” *J. Ilm. Kesehat.*, vol. 3, no. 3, pp. 210–216, 2021, doi: 10.36590/jika.v3i3.202.
- [4] V.A.R.Barao, R.C.Coata, J.A.Shibli, M.Bertolini, and J.G.S.Souza, “Korelasi Glukosa Urine Dan Berat Jenis Urine Pada Penderita Diabetes Melitus Tipe 2 Di RSUD Budhi Asih,” *Braz Dent J.*, vol. 33, no. 1, pp. 1–12, 2022.
- [5] A. M. Argina, “Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes,” *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 29–33, 2020, doi: 10.33096/ijodas.v1i2.11.
- [6] H. Hariawan, A. Fathoni, and D. Purnamawati, “Hubungan Gaya Hidup (Pola Makan dan Aktivitas Fisik) Dengan Kejadian Diabetes Melitus di Rumah Sakit Umum Provinsi NTB,” *J. Keperawatan Terpadu (Integrated Nurs. Journal)*, vol. 1, no. 1, p. 1, 2019, doi: 10.32807/jkt.v1i1.16.
- [7] K. Anis, K. P. A. Nugroho, and C. Natawirarindry, “Gambaran Pola Makan Pra Lansia terkait Risiko Inflamasi Diabetes Melitus Tipe 2 Pada masa Pandemi Covid-19 di Wilayah Kerja Puskesmas Sidorejo Lor, Kota Salatiga,” *J. Sains dan Kesehat.*, vol. 3, no. 6, pp. 807–815, 2021, doi: 10.25026/jsk.v3i6.463.
- [8] Z. Nurushofa, A. Efraim, N. Gori, R. A. Claudia, and D. Melitus, “Edukasi Deteksi Dini Diabetes Melitus Kader Puskesmas Pondok Kacang Timur Tangerang Selatan Pendahuluan Pengabdian masyarakat , partisipasi aktif dosen dapat dilaksanakan dalam berbagai bentuk aplikasi karya dan bakti salah satu bentuk pengabdian mahasiswa,” vol. 5636, no. 1, pp. 86–90, 2023.
- [9] I. M. Karo Karo and H. Hendriyana, “Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score,” *J. Teknol. Terpadu*, vol. 8, no. 2, pp. 94–99, 2022, doi: 10.54914/jtt.v8i2.564.
- [10] G. Abdurrahman, “Jurnal Sistem dan Teknologi Informasi Klasifikasi Penyakit Diabetes Melitus Menggunakan Adaboost Classifier,” *JUSTINDO (Jurnal Sist. dan Teknol. Informasi)*, vol. 7, no. 1, pp. 59–66, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JUSTINDO>
- [11] N. Nurdiana and A. Algifari, “Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus,” *INFOTECHjournal*, vol. 6, no. 2, pp. 18–23, 2020.
- [12] D. Random *et al.*, “Komparasi Performansi Algoritma Pengklasifikasi KNN , Bagging,” pp. 367–372, 2021.
- [13] H. Apriyani and K. Kurniati, “Perbandingan Metode Naive Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus,” *J. Inf. Technol. Ampera*, vol. 1, no. 3, pp. 133–143, 2020, doi: 10.51519/journalita.volume1.iss3.year2020.page133-143.
- [14] N. M. Putry, “Komparasi Algoritma Knn Dan Naive Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus,” *EVOLUSI J. Sains dan Manaj.*, vol. 10, no. 1, 2022, doi: 10.31294/evolusi.v10i1.12514.
- [15] R. G. Whendasmoro and J. Joseph, “Analisis Penerapan Normalisasi Data Dengan Menggunakan Z-Score Pada Kinerja Algoritma K-NN,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 4, p. 872, 2022, doi: 10.30865/jurikom.v9i4.4526.
- [16] S. Mahrani, I. D. Pasi, A. K. Mutmainnah, S. W. P. Samosir, and I. Gunawan, “Proses Pembangunan Smart City Di Indonesia Menggunakan Metode Big Data Analytis Dalam Penerapan E-Commerce,” *Media J. Inform.*, vol. 13, no. 2, p. 57, 2021, doi: 10.35194/mji.v13i2.1866.
- [17] 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.
- [18] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, and E. Suryati Submitted, “Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM),” vol. 4, no. 1, pp. 96–106, 2023, [Online]. Available: <https://doi.org/10.33365/jtsi.v4i1.2445>
- [19] Suhardjono, W. Ganda, and H. Abdul, “Prediksi Kellusan Menggunakan Svm Berbasis Pso,” *Bianglala Inform.*, vol. 7, no. 2, pp. 97–101, 2019.
- [20] R. Susetyoko, W. Yuwono, E. Purwantini, and

- N. Ramadijanti, “Perbandingan Metode Random Forest, Regresi Logistik, Naïve Bayes, dan Multilayer Perceptron Pada Klasifikasi Uang Kuliah Tunggal (UKT),” *J. Infomedia Tek. Inform. Multimed. Jar.*, vol. 7, no. 1, pp. 8–16, 2022.
- [21] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: <http://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- [22] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, “Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation,” *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [23] L. Mardiana, D. Kusnandar, and N. Satyahadewi, “Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak,” *Bimaster Bul. Ilm. Mat. Stat. dan Ter.*, vol. 11, no. 1, pp. 97–102, 2022, [Online]. Available: <https://jurnal.untan.ac.id/index.php/jbmstr/article/view/51608>
- [24] T. Ridwansyah, “Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier,” *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.
- [25] C. B. Sonjaya, A. Fitri, N. Masruriyah, and D. Sulistya, “The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease,” vol. 5, no. 2, pp. 166–175, 2022.

Halaman ini sengaja dikosongkan