

The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease

Chepy Bagustian Sonjaya¹, Anis Fitri Nur Masruriyah², Dwi Sulistya Kusumaningrum³, Adi Rizky Pratama⁴

^{1,2,3,4}Teknik Informatika, Universitas Buana Perjuangan Karawang, Indonesia
if19.chepysonjaya@mhs.ubpkarawang.ac.id

Info Artikel

Sejarah artikel:

Diterima Oktober 2022

Direvisi Desember 2022

Disetujui Desember 2022

Diterbitkan Desember 2022

ABSTRACT

Heart disease in Indonesia, especially in the productive age, there is always an increase in the number of cases. The main cause of the increase in the number of heart patients is an unhealthy lifestyle and diet. The increase in patients with heart disease also has an impact on decreasing the standard of living. With this in mind, there is a need for research related to comparing classification methods on heart disease datasets. The dataset obtained is not balanced so that an oversampling technique is needed. The oversampling technique used is SMOTE. This research method uses Support Vector Machine (SVM) and Logistic Regression (LR). In order for this research method to be applied successfully, the data acquisition, data pre-processing and data transformation techniques are used to ensure accurate results. The model evaluation technique used is K-Fold Cross Validation. Based on the results of the analysis, it showed that the data partition using k-fold cross validation without oversampling gets the same accuracy value but the precision value is quite low. Conversely, if using the SMOTE technique, the accuracy value is as good as the precision value. The results of the SVM accuracy value get a value of 91.69%. LR is 91.76%. While the results of the SVM precision value of 57.81% and LR 54.82%. If using the SVM oversampling technique, the score is 75.79% and the LR is 75.84%. Meanwhile, the precision value obtained in SVM is 75.74%. At LR by 74.77%.

Keywords : Cardiovascular Diseases; Classification; Cross Validation; Logistic Regression; Support Vector Machine.

ABSTRAK

Penyakit jantung di Indonesia terutama pada usia produktif selalu terjadi kenaikan jumlah kasus. Adapun penyebab utama terjadinya kenaikan jumlah pasien jantung adalah gaya hidup dan pola makan yang tidak sehat. Meningkatkan pasien penyakit jantung juga berdampak pada penurunan taraf hidup. Sehingga penelitian ini bertujuan untuk membandingkan metode klasifikasi pada dataset penyakit jantung. Dataset yang didapatkan tidak seimbang sehingga diperlukannya teknik *oversampling*. Teknik *oversampling* yang dipakai yaitu SMOTE. Metode penelitian ini menggunakan *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Agar metode penelitian ini berhasil diterapkan, digunakan teknik akuisisi data, pra-pemrosesan data dan transformasi data agar hasilnya akurat. Teknik evaluasi model yang digunakan yaitu *K-Fold Cross Validation*. Berdasarkan hasil analisis, menunjukkan bahwa partisi data menggunakan *k-fold cross validation* tanpa *oversampling* mendapatkan nilai Akurasi yang sama baiknya namun nilai presisinya cukup rendah. Sebaliknya jika menggunakan teknik SMOTE nilai akurasi sama baiknya dengan nilai presisi. Hasil nilai akurasi SVM mendapatkan nilai sebesar 91,69%. LR sebesar 91.76%. Sementara hasil nilai presisi SVM sebesar 57.81% dan LR 54.82%. Jika menggunakan teknik *oversampling* SVM

mendapatkan nilai sebesar 75.79% dan LR sebesar 75.84%. Sementara hasil nilai presisi yang didapat pada SVM sebesar 75.74%. Pada LR sebesar 74.77%.

Kata Kunci : *Cross Validation*; Klasifikasi; *Logistic Regression*; Penyakit jantung; *Support Vector Machine*.

PENDAHULUAN

Berdasarkan data yang dikumpulkan oleh Kementrian Kesehatan Republik Indonesia [1], penyakit jantung setiap tahunnya mengalami peningkatan dan menempati peringkat tertinggi penyebab kematian di Indonesia terutama pada masa usia produktif. Dijelaskan bahwa penyebab terjadinya kenaikan jumlah kasus pada pasien penyakit jantung dalam suatu populasi suatu waktu disebut juga dengan prevalensi selalu meningkat dikarenakan perubahan gaya hidup yang tidak sehat dan pola makan yang tidak seimbang. Selama masa pandemi COVID-19, pasien yang memiliki penyakit jantung bawaan memiliki peluang keselamatan yang sangat rendah karena dikhawatirkan dapat menyebabkan kondisi pasien semakin memburuk bahkan menyebabkan kematian. Apabila penderita penyakit jantung tidak ditangani dengan baik, maka pasien di usia produktif dapat terancam nyawanya. Sehingga, perlunya pembuktian model aturan klasifikasi yang membantu layanan Kesehatan untuk melakukan penanganan bahkan pencegahan pengidap penyakit jantung pada pasien.

Penyakit jantung biasanya sering digunakan untuk menggambarkan jenis gangguan yang mempengaruhi kinerja jantung [2]. Menurut WHO (Organisasi Kesehatan Dunia) [3] dan CDC [4], penyakit jantung salah satu penyebab utama kematian di Inggris, Amerika Serikat, bahkan hampir di setiap negara. Penyakit atau gangguan jantung sendiri terdiri dari banyak jenis dan macam nama penyakitnya seperti jantung koroner, gangguan irama jantung, penyakit jantung bawaan, kelainan katup, gagal jantung dan serangan jantung [5]. Berdasarkan data WHO [3], terdapat 17.9 juta jiwa meninggal akibat penyakit jantung dari seluruh dunia.

Penelitian yang telah dilakukan sebelumnya yaitu memprediksi tingkat kematian pada pasien Penyakit Jantung Koroner yang mengidap dialisis oleh Mezzatesta et al. [6]. Pada penelitiannya menggunakan algoritma SVM untuk memprediksi dengan *kernel Radial Basis Function* yang di optimasi menggunakan *GridSearch*. Model prediksi yang diuji memiliki hasil akurasi sebesar 95.25%. Kemudian, penelitian yang dilakukan oleh Ghosh et al. [7], pada kasus kardiovaskular menerapkan algoritma ekstrasi fitur RELIEF dan LASSO. Penelitian tersebut memanfaatkan pembelajaran mesin untuk melakukan proses prediksi penyakit kardiovaskular dan mendapatkan hasil akurasi sebesar 99.05%. Selanjutnya, El-Hasnony et al. [8] pada penelitiannya membuat model pencegahan untuk penyakit stroke dan jantung. Penerapan *active learning* pada penelitiannya bertujuan untuk menentukan faktor yang paling berpengaruh pada penyakit stroke dan jantung. Sehingga, berdasarkan faktor-faktor yang mempengaruhi tersebut, tenaga medis dapat melakukan perawatan yang tepat dalam pencegahan penyakit stroke dan jantung.

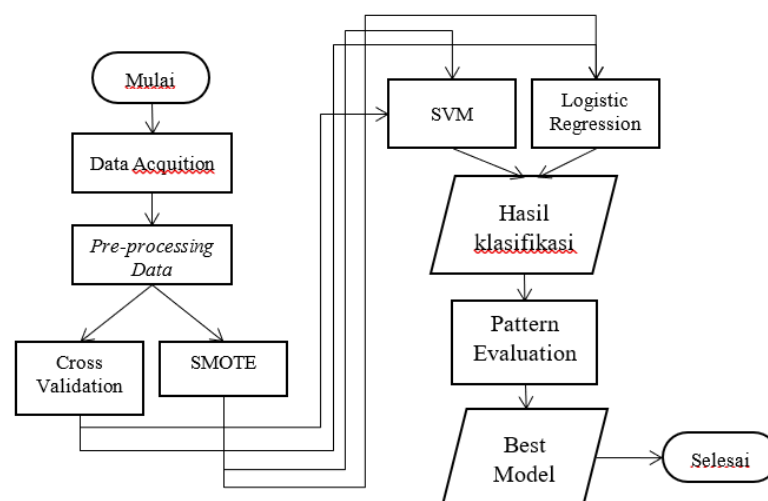
Dalam penelitian ini data yang diperoleh memiliki *class* yang tidak seimbang. Dimana kelas *majority* lebih besar dibandingkan dengan kelas *minority*. *Class imbalanced* dapat mempengaruhi hasil keakuratan klasifikasi. Solusi untuk

mengatasi *imbalanced* data yaitu menggunakan metode *oversampling* SMOTE. Metode *oversampling* pada kelas *minority* dapat membantu mencapai kinerja klasifikasi yang lebih baik. Pada penelitian yang dilakukan oleh Hidayat et al. [9] melakukan perbandingan performa algoritma klasifikasi dari data yang tidak seimbang menggunakan metode SMOTE dan ADASYN. Penerapan metode SMOTE dan ADASYN pada data yang tidak seimbang sangat mempengaruhi performa algoritma SVM, terjadi kenaikan kenaikan pada label *True* (minoritas) pada pengujian pada F1-Score dari 0.597 menjadi 0.815 begitu juga pada presisi dan recall. Namun terjadi penurunan pada label *False* (mayoritas) pada pengujian pada F1-Score dari 0.871 menjadi 0.810 begitu juga pada presisi dan *recall*.

Bersumber dari beberapa referensi penelitian menjelaskan bahwa model klasifikasi mampu membantu tenaga medis dalam menyelesaikan penyakit jantung. Di sisi lain, pada bidang lainnya model klasifikasi juga dapat melakukan prediksi yang baik. Oleh karena itu, penelitian ini bertujuan untuk mengetahui model terbaik dengan menerapkan beberapa algoritma untuk kasus penyakit jantung. Selain itu, penelitian ini juga bertujuan untuk menerapkan teknik ekstraksi fitur untuk mengidentifikasi variabel yang paling mempengaruhi penyakit jantung berdasarkan perhitungan.

METODE

Penelitian ini menggunakan dua metode pembelajaran mesin dalam prediksi penyakit jantung yang terdiri dari *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Dataset yang telah diperoleh dibagi sesuai dengan kriteria kemudian dilakukan pengujian menggunakan dua metode tersebut untuk mendapatkan performa dari masing-masing metode yang digunakan. *Figure 1* Menggambarkan alur pada penelitian ini.



Gambar 1. Diagram Alir Penelitian

Data pasien penyakit jantung dengan total objek yaitu 319.795 data dengan tujuh belas variabel dan satu kelas target. Pada penelitian ini jenis data yang digunakan pada penelitian tugas akhir ini yaitu data Kuantitatif. Data kuantitatif adalah data yang bersifat angka (numerik) yang akan diolah menggunakan metode

statistika. Sedangkan sumber data didapatkan berasal dari catatan medis yang telah di setujui oleh Organisasi Kesehatan Dunia dan dapat diakses pada (2022). Attribut pada penelitian ini yaitu BMI, *Smoking*, *AlcoholDrinking*, *Stroke*, *PhysicalHealth*, *MentalHealth*, *DiffWalking*, *Sex*, *AgeCategory*, *Race*, *Diabetic*, *PhysicalActivity*, *GenHealth*, *SleepTime*, *Asthma*, *KidneyDisease*, *SkinCancer* dan *HeartDisease*.

Penelitian ini menggunakan empat tahap analitika (*data quality analytics*, *descriptive analytics*, *diagnostic analytics* dan *predictive analytics*). Pada tahap pertama pra-pemrosesan data sudah termasuk data *quality analytics* dan *descriptive analytics*. Selanjutnya hasil pra-pemrosesan data diolah untuk mendapatkan hasil *diagnostic analytics* dan *predictive analytics*.

Synthetic Minority Oversampling Technique (SMOTE) diperkenalkan oleh Nithes V. Chawla sebagai salah satu teknik untuk menangani data yang tidak seimbang. Teknik SMOTE digunakan untuk menyeimbangkan jumlah distribusi data pada sampel kelas data minoritas dengan cara menyeleksi data sampel tersebut hingga jumlah sampel data menjadi seimbang dengan jumlah sampel data kelas mayoritas. Metode SMOTE bekerja dengan mengelompokkan data terdekat yang dipilih berdasarkan dari jarak *euclidean* antara kedua data [10].

Support Vector Machine atau lebih dikenal dengan SVM merupakan yang metode digunakan pada masalah regresi dan lebih berfokus pada klasifikasi. Algoritma SVM bisa digunakan untuk melakukan analisis data lalu mengurutkannya ke dalam salah satu dari dua kategori. SVM pada dasarnya membantu menemukan fungsi pemisah atau *hyperplane* yang paling baik memisahkan dua atau lebih kelas di ruang input. *Hyperplanes* dapat berupa garis dalam dua dimensi, dan dapat juga berupa bidang datar di beberapa bidang [11].

Pada algoritma SVM, terdapat 2 kasus dalam memisahkan kelas oleh *Hyperplane* yaitu kelas yang dapat dipisahkan secara sempurna disebut dengan SVM linier dan kelas yang tidak dapat dipisahkan dengan sempurna disebut SVM non linier. Pada dasarnya SVM non linier merupakan solusi dari masalah SVM linier dengan melakukan fungsi *kernel* pada ruang fitur dimensi tinggi [13]. Definisi SVM linier dan nonlinear dapat dilihat tabel 1.

Tabel 1. SVM Linier dan non Linear

SVM	Jenis Kernel	Definisi Rumus
Linier	Linier	$K(x,y) = x.y$
	<i>Polynomial</i>	$K(x,y) = (x.y + 1)^p$
Non linier	RBF	$K(x,y) = e^{- x.y ^2/2\sigma^2}$
	<i>Sigmoid</i>	$(x,y) = \tanh(Kx.y - \delta)$

Logistic Regression merupakan algoritma terbimbing pada masalah regresi dan klasifikasi. Algoritma ini menggunakan probabilitas untuk memprediksi klasifikasi data kategorikal. Jenis analisisnya dapat digunakan jika variabel dependen adalah variabel dikotomi (dua opsi). Variabel dikotomi biasanya hanya terdiri dari dua nilai, yang mewakili terjadinya atau tidak adanya suatu peristiwa yang biasanya diberi angka 0 atau 1. Algoritma ini juga dikenal sebagai regresi

logistik, pengklasifikasi log-linear atau klasifikasi entropi maksimum (MaxEnt) [13].

$$p = \frac{1}{1+e^{-(a+bX)}} \quad (1)$$

Pada persamaan (1) p adalah probabilitas dan a, b akan menjadi parameter model, X adalah faktor. Evaluasi dilakukan untuk memilih dataset dan metode klasifikasi yang dapat mencapai tingkat akurasi yang tinggi. Evaluasi dalam penelitian adalah memperhatikan *confussion matrix*. *Confussion Matrix* adalah alat untuk menentukan sejauh mana pengklasifikasi dapat mengenali atau memprediksi kelas dalam data [15]. *Confussion Matrix* dilakukan dengan model yang bertujuan untuk mengetahui seberapa baik performa model dengan menggunakan data uji. Evaluasi didasarkan presisi dan akurasi. Perhitungan tersebut disebut dengan *confussion matriks* [16]. Perhitungan akurasi dan presisi ditunjukkan pada tabel 2 dan seperti yang dijelaskan pada persamaan 2 dan 3.

Tabel 2. Confussion Matrix

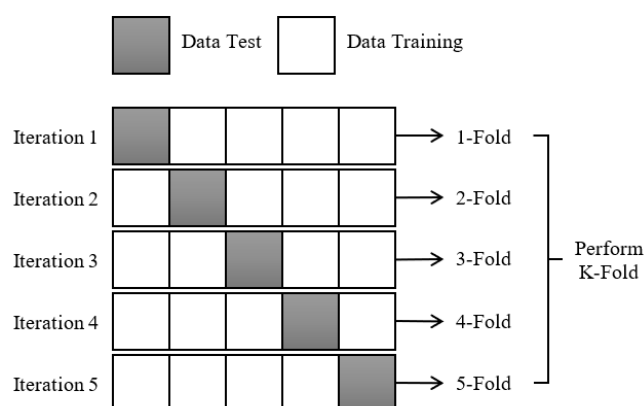
Prediksi	Aktual	
	A	B
A	TP (True Positive)	FP (False Positive)
B	FN (False Negative)	TN (True Negative)

$$\text{Akurasi} = \frac{TP+TN}{TP + TN+FP+FN} \quad (2)$$

$$\text{Presisi} = \frac{\text{True Positive}}{\text{True Positive} + \text{False positive}} \quad (3)$$

Pada persamaan (2) akurasi bertujuan untuk memprediksi *true positive* dan *true negative* dari keseluruhan data. Akurasi ini memiliki pernyataan bahwa seberapa persen pasien yang diprediksi terkena atau tidaknya penyakit jantung. Persamaan (3) presisi bertujuan untuk melakukan perbandingan antara *true positive* dengan banyaknya data yang juga di prediksi positif. Presisi ini memiliki pernyataan bahwa seberapa persen pasien yang terkena penyakit jantung dari seluruh pasien yang juga terkena penyakit jantung.

Selanjutnya, Pengujian pada penelitian dilakukan evaluasi menggunakan Teknik *K-Fold Cross Validation*. Pengerjaan dari Teknik ini yaitu dengan membagi data menjadi data uji dan data latih sebanyak K. Gambaran Teknik *K-Fold Cross Validation* pada gambar 2.



Gambar 2. Gambaran K-Fold Cross Validation
Sumber : Ansari [17]

HASIL DAN PEMBAHASAN

Pra-pemrosesan Data

Penelitian ini dataset yang digunakan bersumber dari catatan medis yang telah di setujui oleh Organisasi Kesehatan Dunia yang diakses pada tahun 2022 dengan data berjumlah 319.795 dengan jumlah atribut 17. Data penyakit jantung yang diperoleh tidak seimbang, maka perlu melakukan Teknik oversampling, sehingga data tidak seimbang bisa menjadi seimbang. Teknik oversampling yang digunakan yaitu SMOTE.

Data dibersihkan agar tidak terdapat data duplikat ataupun bernilai *null*. Data yang berjenis kategori ditransformasikan menjadi angka dengan 1 = *yes/male*, lalu 0 = *no/female*. Tujuannya agar mesin dapat melakukan pengujian, karena mesin hanya dapat membaca numerik untuk melakukan pengujian. Sedangkan data yang berjenis numerik dilakukan normalisasi dengan bertujuan untuk proses penskalaan data sehingga data terletak pada rentang tertentu.

Modelling

Pada tahap *modelling* akan dilakukan klasifikasi terhadap data. Sebelum dilakukan klasifikasi pada penelitian ini data terlebih dahulu dipartisi data uji dan data latih menggunakan teknik *k-fold cross validation*. Kemudian, dataset yang telah melewati tahapan *pre-processing* dilanjutkan dengan proses pembelajaran menggunakan metode klasifikasi supervised learning yaitu SVM dan *Logistic Regression*. Sehingga hasilnya nanti bisa dibandingkan sebagai data terbaik.

Penelitian ini menggunakan metode SVM dan Logistic Regression sebagai model klasifikasi dan menggunakan SMOTE sebagai Teknik oversampling yang akan digunakan. Selanjutnya, performa akan dibandingkan menggunakan algoritma klasifikasi SVM dan Logistic Regression tanpa oversampling dengan menggunakan Teknik oversampling.

Evaluasi

Partisi data *k-fold cross validation* digunakan pada pengujian karena metode ini dapat digunakan untuk mengurangi bias yang terdapat pada data *random*. Dalam penelitian ini digunakan *10-fold cross validation* dengan membagi dataset

menjadi 10 *fold* yang berbeda dengan ukuran yang sama dan melakukan pelatihan dan pengujian pada model sebanyak 10 kali.

Selanjutnya dilakukan klasifikasi dengan algoritma *Support Vector Machine* dan *Logistic Regression* tanpa *oversampling*. Dari setiap metode klasifikasi yang digunakan didapat nilai akurasi dan presisi. Berikut hasil *k-fold* 1 sampai *k-fold* 10 yang didapatkan pada table 3:

Tabel 3. Hasil Akurasi dan presisi K-fold 10 tanpa *oversampling* (SMOTE)

K-Fold	SVM		LR	
	Akurasi	Presisi	Akurasi	Presisi
1	91.61%	62.01%	91.60%	54.06%
2	91.76%	55.44%	91.74%	51.45%
3	91.50%	60.09%	91.58%	55.99%
4	91.38%	62.5%	91.48%	57.76%
5	91.89%	64.48%	91.79%	52.26%
6	91.46%	56.66%	91.46%	52.19%
7	91.30%	57.14%	91.32%	53.25%
8	91.50%	65.14%	91.60%	57.60%
9	91.53%	63.29%	91.49%	53.47%
10	91.69%	57.81%	91.76%	54.82%

Berdasarkan table 3, diperoleh bahwa akurasi tertinggi terdapat pada algoritma SVM dan LR di *K-fold* 5 dengan akurasi SVM 91.89% dan LR 91.79%. Sedangkan, pada presisi tertinggi terdapat pada *K-fold* 4 dengan algoritma LR nilai presisi nya yaitu 57.76%. Selanjutnya dilakukan klasifikasi dengan algoritma *Support Vector Machine* dan *Logistic Regression* dengan *oversampling*, berikut hasil *k-fold* 1 sampai *k-fold* 10 yang didapatkan pada tabel 4.

Tabel 4. Hasil Akurasi dan presisi K-fold 10 menggunakan *oversampling* (SMOTE)

K-Fold	SVM + SMOTE		LR + SMOTE	
	Akurasi	Presisi	Akurasi	Presisi
1	75.82%	74.64%	75.81%	74.80%
2	75.84%	74.61%	75.83%	74.79%
3	75.56%	74.27%	75.55%	74.52%
4	76.04%	74.87%	76.03%	75.11%
5	75.69%	74.45%	75.68%	74.60%
6	75.90%	74.39%	75.91%	74.59%
7	75.61%	74.00%	75.62%	74.20%
8	75.84%	74.72%	75.82%	74.89%
9	75.75%	74.52%	75.79%	74.81%
10	75.79%	74.54%	75.84%	74.77%

Berdasarkan tabel 4 diperoleh bahwa nilai dan presisi yang didapatkan mengalami sedikit perubahan apabila menggunakan teknik *oversampling*. Pada nilai akurasi mengalami penurunan, tetapi nilai presisi menjadi semakin meningkat, sehingga akurasi dan presisi mendapatkan nilai yang mendekati dan stabil. Pada nilai akurasi terbesar yaitu algoritma SVM pada *k-fold* 4 dengan nilai akurasi yaitu 76.04%, nilai presisi terbesar yaitu algoritma *Logistic Regression* pada *k-fold* 4 dengan nilai presisi terbesar yaitu 75.11%.

Perbandingan Model

Berdasarkan pada penelitian sebelumnya [16] *K-Fold* yang diambil sebagai perbandingan yaitu *K-Fold* 10, karena banyak beberapa penelitian menunjukkan bahwa *K-Fold* 10 merupakan pilihan terbaik untuk mendapatkan estimasi yang akurat. Maka hasil evaluasi yang sudah didapat, dibandingkan pada tabel 5.

Tabel 5. Perbandingan Performa

<i>K-Fold Cross Validation</i>		
Metode	Akurasi	Presisi
SVM	91.69%	57.81%
LR	91.76%	54.82%
SVM + SMOTE	75.79%	74.54%
LR + SMOTE	75.84%	74.77%

Pada hasil penelitian menggunakan *K-Fold Cross Validation* di dapatkan nilai presisi lebih rendah dibandingkan dengan akurasi. Maka dari kedua model algoritma klasifikasi yaitu SVM dan *Logistic Regression* meskipun hasil perulangan dari setiap percobaan tidak selalu tepat atau kurang presisi, namun klasifikasi yang dibuat menggunakan model tersebut terbukti mendapatkan hasil sangat akurat.

Jika dilakukan proses *oversampling* menggunakan Teknik SMOTE, presisi yang dihasilkan mendapat kenaikan dibandingkan dengan tidak melakukan teknik *oversampling*. Namun, akurasi yang didapat mengalami pengurangan yang cukup besar juga, yang dimana hasil akurasi dan presisi menjadi lebih stabil dibandingkan tidak menggunakan teknik *oversampling*.

Pada tabel 5 perbandingan tanpa *oversampling* menggunakan *K-Fold Cross Validation Logistic regression* lebih unggul dibandingkan dengan SVM. *Logistic regression* mendapatkan nilai akurasi sebesar 91.76% dan presisi 54.82%. Sedangkan, SVM mendapatkan nilai akurasi sebesar 91.69% dan presisi 57.81%. Perbandingan dengan teknik *oversampling Logistic regression* tetap lebih unggul dibandingkan dengan SVM. *Logistic regression* mendapatkan nilai akurasi sebesar 78.84% dan presisi 74.77%. Sedangkan, SVM mendapatkan nilai akurasi sebesar 75.79% dan presisi 74.54%.

PENUTUP

Berdasarkan hasil penelitian yang diperoleh, nilai akurasi tertinggi yaitu membandingkan tanpa *oversampling* menggunakan *K-Fold Cross Validation Logistic regression* dengan hasil lebih unggul dibandingkan dengan SVM. *Logistic regression* mendapatkan nilai akurasi sebesar 91.76% dan presisi bernilai 54.82%. Sedangkan, SVM mencapai nilai akurasi sebesar 91.69% dan nilai presisi 57.81%. Perbandingan dengan teknik *oversampling Logistic regression* secara konsisten mengungguli SVM. *Logistic regression* mendapatkan nilai akurasi sebesar 78.84% dan nilai presisi 74.77%. Sedangkan, SVM mendapatkan nilai akurasi sebesar 75.79% dan nilai presisi 74.54%.

Berdasarkan Penelitian yang sudah dilakukan hanya membandingkan dua algoritma klasifikasi yaitu *Support Vector Machine* (SVM) dan *Logistic Regression* (LR). Pada penelitian selanjutnya mungkin bisa digunakan algoritma klasifikasi

yang lainnya, sehingga mendapatkan performa yang lebih baik menggunakan dataset penyakit jantung.

DAFTAR PUSTAKA

- [1] "Kementerian Kesehatan Republik Indonesia," 2021. <https://www.kemkes.go.id/article/view/21093000002/penyakit-jantung-koroner-didominasi-masyarakat-kota.html> (accessed Aug. 04, 2022).
- [2] 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>
- [3] "Cardiovascular diseases (CVDs)." [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)) (accessed Oct. 02, 2022).
- [4] C. W. Tsao *et al.*, "Heart Disease and Stroke Statistics-2022 Update: A Report from the American Heart Association," *Circulation*, vol. 145, no. 8, pp. E153–E639, Feb. 2022, doi: 10.1161/CIR.0000000000001052.
- [5] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [6] S. Mezzatesta, C. Torino, P. De Meo, G. Fiumara, and A. Vilasi, "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis," *Comput. Methods Programs Biomed.*, vol. 177, pp. 9–15, Aug. 2019, doi: 10.1016/j.cmpb.2019.05.005.
- [7] P. Ghosh *et al.*, "Efficient prediction of cardiovascular disease using machine learning algorithms with relief and lasso feature selection techniques," *IEEE Access*, vol. 9, pp. 19304–19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
- [8] I. M. El-Hasnony, O. M. Elzeki, A. Alshehri, and H. Salem, "Multi-Label Active Learning-Based Machine Learning Model for Heart Disease Prediction," *Sensors*, vol. 22, no. 3, 2022, doi: 10.3390/s22031184.
- [9] W. Hidayat, M. Ardiansyah, and A. Setyanto, "Pengaruh Algoritma ADASYN dan SMOTE terhadap Performa Support Vector Machine pada Ketidakseimbangan Dataset Airbnb," *Edumatic J. Pendidik. Inform.*, vol. 5, no. 1, pp. 11–20, 2021, doi: 10.29408/edumatic.v5i1.3125.
- [10] R. A. Nurdian, Mujib Ridwan, and Ahmad Yusuf, "Komparasi Metode SMOTE dan ADASYN dalam Meningkatkan Performa Klasifikasi Herregistrasi Mahasiswa Baru," *J. Tek. Inform. dan Sist. Inf.*, vol. 8, no. 1, 2022, doi: 10.28932/jutisi.v8i1.4004.
- [11] Trivusi, "Penjelasan Lengkap Algoritma Support Vector Machine (SVM) - Trivusi," 2022. <https://www.trivusi.web.id/2022/04/algoritma-svm.html> (accessed Aug. 06, 2022).
- [12] M. Aminullah, *Perbandingan Performa Klasifikasi Machine Learning dengan Teknik Resampling pada Dataset Tidak Seimbang*. 2021. [Online]. Available: <https://repository.uinjkt.ac.id/dspace/bitstream/123456789/57648/1/MUHAMMAD AMINULLAH-FST.pdf>
- [13] F. Handayani, "Komparasi Support Vector Machine, Logistic Regression Dan Artificial Neural Network Dalam Prediksi Penyakit Jantung," *J. Edukasi dan*

- Penelit. Inform.*, vol. 7, no. 3, p. 329, 2021, doi: 10.26418/jp.v7i3.48053.
- [14] W. Willy, D. P. Rini, and S. Samsuryadi, "Perbandingan Algoritma Random Forest Classifier, Support Vector Machine dan Logistic Regression Clasifier Pada Masalah High Dimension (Studi Kasus: Klasifikasi Fake News)," *J. Media Inform. Budidarma*, vol. 5, no. 4, p. 1720, 2021, doi: 10.30865/mib.v5i4.3177.
- [15] P. R. Sihombing, "Perbandingan Metode Artificial Neural Network (ANN) dan Support Vector Machine (SVM) untuk Klasifikasi Kinerja Perusahaan Daerah Air Minum (PDAM) di Indonesia," *J. Ilmu Komput.*, vol. 13, no. 1, p. 9, 2020, doi: 10.24843/jik.2020.v13.i01.p02.
- [16] U. Amelia *et al.*, "Implementasi Algoritma Support Vector Machine (Svm) Untuk Prediksi Penyakit Stroke Dengan Atribut Berpengaruh," vol. III, pp. 254–259, 2022.
- [17] F. Ansari, "Cross Validation Techniques. This write-up contains explanation of... | by Faizan Ansari | Analytics Vidhya | Medium," *Medium.com*, 2021. <https://medium.com/analytics-vidhya/cross-validation-techniques-bacb582097bc> (accessed Aug. 31, 2022).