

IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI PENYAKIT AIDS.

Sonia Salfira¹, M. Chaerul Akbar², Muhammad Salim³

Sonia Salfira mengerjakan Rapidminer dan edit semua Jurnal, M. Chaerul Akbar mengerjakan Pendahuluan,

Muhammad Salim mengerjakan bagian Abstrak

Program Studi Teknik dan Ilmu Komputer S1, Fakultas Ilmu Komputer

cUniversitas Muhammadiyah Bima

Jalan Anggrek No.16 Ranggo Nae Kec. Rasanae Barat Kota Bima.

Website: www.umbima.ac.id Email: ilkom.umbima@gmail.com

Abstrak : Menurut World Health Organization (WHO), AIDS (Acquired Immunodeficiency Syndrome) merupakan penyakit menular dari satu orang ke orang lain yang disebabkan oleh virus HIV (Human Immunodeficiency Virus). AIDS adalah tahap akhir dari infeksi HIV, merusak sistem kekebalan tubuh sampai tidak lagi berfungsi, menyebabkan sekumpulan gejala penyakit mematikan. Virus HIV adalah penyakit seumur hidup yang tinggal di dalam tubuh individu yang terinfeksi. Pada penelitian ini telah menggunakan data penyakit AIDS dengan atribut usia, berat badan, ras, jenis kelamin, aktivitas seksual, hemofilia, riwayat penggunaan obat IV dan jumlah CD4/CD8. Maka dilakukan klasifikasi pada data Penyakit AIDS dengan implementasi algoritma Decision Tree yang unggul dan menghasilkan tingkat akurasi yang tinggi. Hasil klasifikasi dengan tools RapidMiner yang dilakukan dengan penerapan algoritma Decision Tree maka diperoleh nilai akurasi 83.62%, rata-rata recall 91.53%, rata-rata precision 74.45 %, nilai MSE (Mean Squared Error) 0.131 +/- 0.273, nilai RMSE (Root Mean Squared Error) 0.362 +/- 0.000 dan nilai F1-Score 68.46%. Sehingga disimpulkan bahwa kinerja Decision Tree dalam mengklasifikasi data penyakit AIDS cukup tinggi.

Kata Kunci: Data Mining; Klasifikasi; Decision Tree; Penyakit AIDS; RapidMiner

Abstract : According to the World Health Organization (WHO), AIDS (Acquired Immunodeficiency Syndrome) is a contagious disease from one person to another caused by the HIV virus. (Human Immunodeficiency Virus). AIDS is the final stage of HIV infection, damaging the immune system until it no longer functions, causing a set of deadly disease symptoms. The HIV virus is a lifelong disease that lives within the body of an infected individual. The study used AIDS disease data with attributes of age, weight, race, gender, sexual activity, hemophilia, history of IV drug use and CD4/CD8 count. The classification with the tools of RapidMiner was done with the application of the Decision Tree algorithm and obtained an accuracy of 83.62%, a recall average of 91.53%, an average precision of 74.45 %, MSE (Mean Squared Error) 0.131 +/- 0.273, RMSE (Root Mean Squared Error) 0.362 +/- 0.000, and F1-Score 68.46%. So it was concluded that the performance of decision tree in classifying chronic AIDS data was fairly good.

Keywords: Data Mining; Classification; Decision Tree; AIDS; RapidMiner

1. PENDAHULUAN

Pola hidup yang sehat adalah harapan semua orang, namun kenyataannya banyak yang menjalani pola hidup tidak sehat. Pola hidup yang tidak sehat ini dapat menyebabkan berbagai penyakit menular yang dapat menurunkan kualitas hidup seseorang. Kualitas hidup memungkinkan seseorang menikmati kehidupannya, yang sangat bergantung pada kesehatan fisik dan mental.

Menurut World Health Organization (WHO) mendefinisikan AIDS (Acquired Immunodeficiency Syndrome) adalah penyakit menular dari satu orang ke orang lain yang disebabkan oleh virus HIV (Human Immunodeficiency Virus). AIDS adalah tahap akhir dari infeksi HIV, merusak sistem kekebalan tubuh sampai tidak lagi berfungsi, menyebabkan sekumpulan gejala penyakit mematikan. [1]. Virus HIV adalah penyakit seumur hidup yang tinggal di dalam tubuh individu yang terinfeksi [2]

Berdasarkan data dari Kementerian Kesehatan RI, jumlah penderita HIV/AIDS di Indonesia terus meningkat setiap tahun. Data dari Kementerian Kesehatan mengungkapkan bahwa mayoritas kasus terkonsentrasi di daerah-daerah tertentu, dengan DKI Jakarta memiliki jumlah pasien HIV+ tertinggi. Provinsi lain dengan kasus yang signifikan termasuk Jawa Timur, Jawa Barat, Papua, dan Jawa Tengah. Data juga menunjukkan bahwa kelompok usia yang paling terpengaruh oleh HIV+ adalah antara 24-49 tahun dan 20-24 tahun. [3]

Dengan meningkatnya jumlah penderita HIV/AIDS, diperlukan upaya pencegahan dan pengobatan yang tepat untuk mengendalikan penyebaran penyakit ini. Pencegahan dapat dilakukan melalui edukasi tentang praktik seksual yang aman, penyediaan jarum suntik steril, dan program pencegahan penularan dari ibu ke anak. Pengobatan untuk HIV/AIDS meliputi terapi antiretroviral (ART) yang membantu menekan replikasi virus dan memperlambat perkembangan penyakit. Data mining dengan Decision Tree dapat digunakan untuk mengolah data epidemiologi guna memprediksi penyebaran HIV/AIDS dan merancang intervensi kesehatan yang lebih efektif.

Data Mining adalah bidang multidimensional yang terdiri dari berbagai aspek, misalnya, pembelajaran, statistik, teknologi informasi, kecerdasan buatan, pengambilan dan visualisasi data. [4]. Algoritma Decision Tree (Pohon Keputusan) merupakan algoritma pembelajaran yang diawasi yang umum digunakan untuk masalah klasifikasi. Pohon Keputusan dapat mengatur kelas pada tingkat yang akurat, menjadikannya sempurna untuk tugas klasifikasi [5]. Penelitian terdahulu sudah meneliti terhadap kelulusan mahasiswa telah banyak dilakukan diantaranya yaitu “Prediksi Kelulusan Mahasiswa Diploma dengan Komparasi Algoritma Klasifikasi”. Penelitian ini mengkomparasi 5 algoritma data mining yaitu Decision tree C4.5, Naive Bayes, K-NN, rule Induction, dan random forest dengan tujuan untuk menentukan metode yang paling akurat. Data yang digunakan yaitu data mahasiswa AMIK BSI Pontianak angkatan 2013-2017 prodi Informatika sebanyak 394 record. Hasil penelitian menunjukkan bahwa algoritma yang paling optimal adalah Decision tree C4.5 dengan akurasi sebesar 90,85% [6]

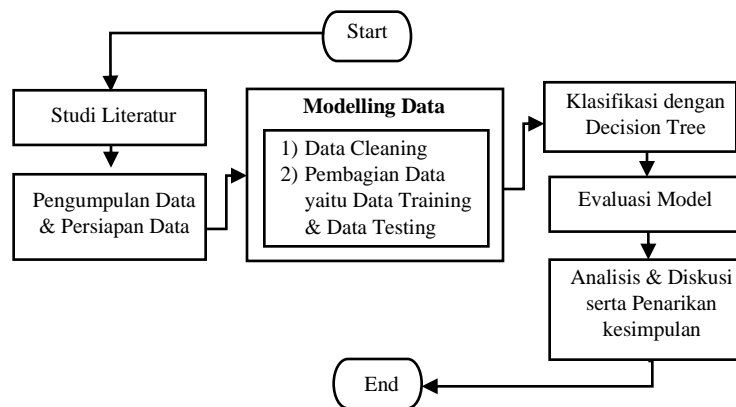
Rakhman pada tahun 2017 telah melakukan studi tentang memprediksi akurasi kelulusan siswa menggunakan metode Decision Tree berdasarkan Particle Swarm Optimization (PSO). Penelitian ini bertujuan untuk menghindari siswa yang lulus terlambat. Hasil yang didapat adalah Algoritma Decision Tree menghasilkan nilai akurasi sebesar 96,57 % dan nilai AUC sebesar 0.942 dengan tingkat diagnosa Excellent Classification, namun setelah dilakukan penambahan yaitu Algoritma Decision Tree berbasis Particle Swarm Optimization nilai akurasi sebesar 97.19 % dan nilai AUC sebesar 0.969 dengan tingkat diagnosa Excellent Classification. Sehingga kedua metode tersebut memiliki perbedaan tingkat akurasi yaitu sebesar 0.62 dan perbedaan nilai AUC sebesar 0,027 [7].

Demikian pula, Rohmawan mengeksplorasi memprediksi kelulusan siswa tepat waktu menggunakan metode Decision Tree dan Jaringan Neural Buatan. Penelitian ini bertujuan untuk meningkatkan akurasi prediksi lulusan dengan menggunakan metode yang berbeda [8] . Selain itu, terdapat penelitian mengenai Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita untuk dapat mengetahui status gizi pada balita. Penelitian ini menggunakan data sebanyak 123 data dengan kelahiran balita dari 2015 sampai 2019. Tujuan penelitian ini untuk mengetahui pohon keputusan berdasarkan data yang ada untuk dianalisa kesesuaiannya dan dilakukan klasifikasi. Hasil uji coba menunjukkan terdapat kesesuaian Decision Tree dengan data yang ada. Algoritma C4.5 memberikan akurasi yang lebih baik karena algoritma C4.5 membangun pohon dengan jumlah cabang tiap simpul sesuai dengan nilai simpul tersebut. Dengan demikian dapat disimpulkan bahwa algoritma C4.5 mampu mengklasifikasikan status gizi balita [9].

Berdasarkan hasil pembahasan yang telah dipaparkan pada pendahuluan maka akan dilakukan klasifikasi terhadap pasien yang mengidap AIDS. Oleh karena itu pada penelitian ini mengangkat judul yaitu tentang Implementasi Algoritma Decision Tree Untuk Klasifikasi Penyakit AIDS.

2. METODE PENELITIAN

Penelitian ini menerapkan metode riset eksperimental yang terdiri atas tahapan (1) Studi Literatur, (2) Pengumpulan Data & Persiapan Data (3) Modelling Data, (4) Klasifikasi Dengan Decision Tree, (5) Evaluasi Model, (8) Analisis dan Diskusi serta Penarikan Kesimpulan. Secara rinci ditunjukkan pada **Gambar 2.1**.



Gambar 2.1. Alur Kerja Penelitian

2.1. Studi Literatur

Pada tahap mempelajari literature di penelitian ini, literatur yang dipakai adalah berupa artikel dan paper penelitian terdahulu di google scholar yang didapatkan sebelumnya untuk menemukan referensi judul yang tentu saja berhubungan dengan metode Klasifikasi Decision Tree sebagai pendukung.

2.2. Pengumpulan Data & Persiapan Data

Data dikumpulkan dari website *Kaggle : Your home for datascience*, data set AIDS Virus Infection Prediction. Setelah data terkumpul, Saya mempersiapkan data tersebut dengan cara menganalisisnya untuk mengetahui isi dari data tersebut dan untuk memperbaiki atau merapikan data jika terdapat data yang berantakan.

2.3. Modelling Data

Dengan pengolahan data preprocessing yang baik ini dapat mempengaruhi hasil dari data mining yang mana dampaknya akan lebih menghasilkan akurasi yang lebih baik dan memproses informasi baru yang lebih jelas, tahapan ini terdiri dari yaitu data cleansing, dan data pembagian data menjadi 2 yaitu data training (Latihan) dan data testing (Test). [10]

2.4. Klasifikasi dengan Naïve Bayes Classifier (NBC) dan Decision Tree

2.4.1. Klasifikasi

Klasifikasi dapat didefinisikan selaku pekerjaan yang melaksanakan pelatihan/pembelajaran terhadap fungsi target yang memetakan tiap set fitur (atribute) ke satu jumlah label kelas yang ada. Klasifikasi ialah suatu buat memperhitungkan objek informasi buat memasukkannya kedalam kelas tertentu dari jumlah kelas yang ada. Klasifikasi menerapkan pembangunan model bersumber pada informasi latih dan data testing untuk pengujian.

2.4.2. Decision Tree

Algoritma Decision Tree adalah model prediktif yang dapat digunakan untuk mewakili model pengklasifikasi dan regresi. Dalam penelitian, operasi Decision Tree mengacu pada model hirarki keputusan. Pohon klasifikasi sering digunakan dalam bidang terapan seperti keuangan, pemasaran, teknik dan kedokteran. Selain itu Decision Tree merupakan supervised learning algorithm yang mampu menganalisis data numerik dan kategorikal.

2.5. Evaluasi Model

Evaluasi model adalah tahap di mana kinerja model yang telah dibangun diuji untuk memastikan bahwa ia bekerja dengan baik dan sesuai dengan tujuan penelitian.

2.6. Analisis dan Diskusi serta Penarikan Kesimpulan

Tahap ini melibatkan presentasi hasil dari analisis data dari Preprocessing Modelling Data sampai dengan pengujian akurasi dan pembahasan tentang implikasi hasil tersebut terhadap pertanyaan penelitian dan literatur yang ada. Lalu tahap terakhir yaitu tahap penarikan kesimpulan, tahap ini tentu saja sangat berkaitan erat dengan tahapan sebelumnya terutama pada tahap pengujian akurasi yang didapatkan sebelumnya [11].

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Penelitian ini menggunakan pengumpulan data penderita AIDS dari *Website Kaggle: Your Home Datascience*. Data menunjukkan bahwa atribut yang mempengaruhi AIDS adalah usia, berat badan, ras, jenis kelamin, aktivitas seksual, hemofilia, riwayat penggunaan obat IV dan jumlah CD4/CD8. AIDS adalah penyakit tahunan yang diderita akibat buruknya pola hidup, oleh karena itu pada penelitian ini dilakukan klasifikasi atribut data untuk menentukan nilai akurasi dalam mengidentifikasi AIDS.

No.	Waktu	Jenis pengobatan (trt)	Umur	Berat badan (kg)	Hemo	Homo	...	Infected
1.	948	2	48	90	0	0	...	tidak
2.	1002	3	61	49	0	0	...	ya
3.	961	3	45	88	0	1	...	tidak
4.	1166	3	47	85	0	1	...	tidak
...
2139	1045	3	45	77	1	0	...	tidak

Tabel 3.1. Pengumpulan Dataset

3.2. Preprocessing Modelling Data

Proses selanjutnya dilakukan data cleaning, data Cleaning adalah proses membersihkan data dari data noise [12]. Selain itu, Data cleaning juga akan membersihkan atribut-atribut yang tidak relevan. Data mining juga akan mempengaruhi hasil informasi dari tekning data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya [13].

Tahap preprocessing dimulai dengan 2139 record data. Setelah dilakukan proses cleaning, tidak ada pengurangan jumlah data karena semua data lengkap tanpa missing values.

Name	Type	Missing	Statistics		Filter (26 / 26 attributes):
Waktu	Integer	0	Min	Max	Average
			62	1231	888.643
Jenis pengobatan (trt)	Integer	0	Min	Max	Average
			0	3	1.460
Umur	Integer	0	Min	Max	Average
			12	70	35.395
Berat badan (kg)	Integer	0	Min	Max	Average
			41	127	75.890
hemo	Integer	0	Min	Max	Average
			0	1	0.098
homo	Integer	0	Min	Max	Average
			0	1	0.654
Riwayat penggunaan obat IV	Integer	0	Min	Max	Average
			0	1	0.157

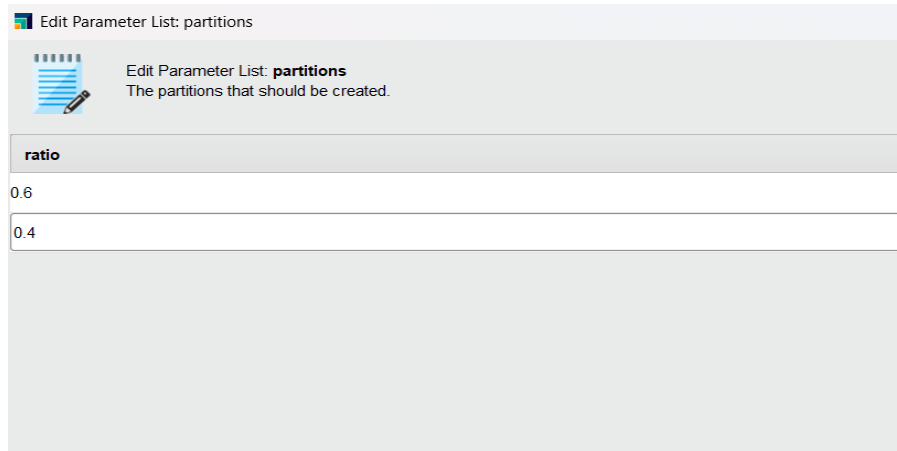
Showing attributes 1 - 26 Examples: 428 Special Attributes: 4 Regular Attributes: 22

Gambar 3.2. Statistics Data

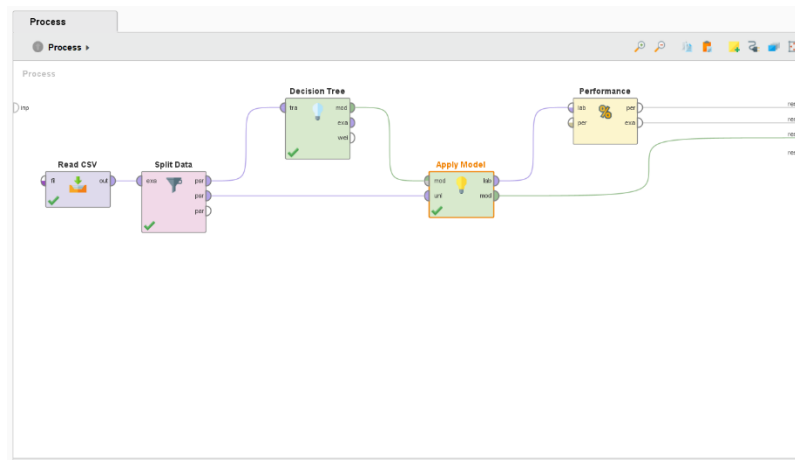
Gambar 3.2 diatas menunjukkan bahwa tidak terdapat missing value. Setelah data selesai melalui proses cleaning, tahap berikutnya adalah membagi data tersebut menjadi dua bagian yaitu data training dan data testing. Data training dan data testing merupakan dua jenis data yang penting dalam proses data mining.

Data training digunakan sebagai data rujukan dalam perhitungan algoritma Decion Tree, sedangkan Data Testing digunakan untuk mengukur sejauh mana algoritma Decion Tree berhasil melakukan klasifikasi dengan benar. Oleh karena itu, data yang ada pada testing set seharusnya tidak boleh ada pada training set sehingga dapat diketahui apakah model classifier sudah “pintar” dalam melakukan klasifikasi [14]. Proses data testing dan training pada machine learning melibatkan pengujian keakuratan model pada dataset yang telah dipisahkan [15]. Dalam implementasi algoritma klasifikasi, evaluasi perbandingan antara data training dan data testing diperlukan untuk menentukan keefektifan model yang dibuat.

Dalam penelitian ini, Dataset proses model Algoritma Decision Tree pada tahap training dibagi menjadi dua partisi yaitu 60% untuk data training dan 40% untuk data testing Dalam *Tools RapidMiner*, langkah ini disebut sebagai pembagian dataset untuk mempersiapkan analisis data mining yang lebih komprehensif (mendalam). seperti yang terlihat pada gambar visualisasi pembagian data terlihat pada **gambar 3.3**.



Gambar 3.3. Split Data Training dan Testing Decision Tree



Gambar 3.4. Proses Implementasi Model Rapid Miner Decision Tree AIDS

3.3. Evaluasi Model Hasil dari Implementasi Algoritma Decision Tree

	true tidak	true ya	class precision
pred. tidak	378	59	86.50%
pred. ya	35	102	74.45%
class recall	91.53%	63.35%	

Tabel 3.3. Confussion Matrix Decision Tree

Confusion matrix yang ditunjukkan oleh **tabel 3.3** diatas adalah alat yang digunakan untuk mengevaluasi model klasifikasi dengan menentukan prediksi memberi informasi hasil nilai true positive, dan nilai true negative. Matrix ini membandingkan hasil prediksi dengan kelas asli dari data input, sehingga berisi informasi mengenai nilai aktual dan prediksi pada klasifikasi.

Hasil pengujian Confussion Matrix dari dua kelas status pasien yang dapat dianalisis dengan bentuk hasil rata-rata precision, rata-rata recall, dan accuracy [16].

Parameter	Nilai
Precision	74.45 %
Recall	91.91 %
Accuracy	83.62%
MSE (Mean Squared Error)	0.131 +/- 0.273
RMSE (Root Mean Squared Error)	0.362 +/- 0.000
F1-Score	68.46 %

Tabel 3.4. Hasil Data Pengujian

Berdasarkan dari tabel 3.4 diatas hasil klasifikasi menggunakan tools RapidMiner dengan algoritma Decision Tree diperoleh nilai akurasi sebesar 83.62 %.

4. KESIMPULAN

Data Penyakit AIDS yang dikumpulkan dari website Kaggle sebanyak 2139 record data, melalui preprocessing data di bagi menjadi dua bagian data berupa 60% data training dan 40% data testing. Klasifikasi yang dilakukan pada penelitian ini menggunakan tools RapidMiner dengan implementasi algoritma Decision Tree. Validasi dalam menguji nilai akurasi maka diperoleh hasil bahwa algoritma Decision Tree memiliki akurasi yang cukup tinggi dengan presentase 83.62%.

Dari hasil ini, dapat disimpulkan bahwa model Decision Tree memiliki performa yang baik dalam mengklasifikasikan data penyakit AIDS, dengan tingkat akurasi yang tinggi dan nilai recall yang sangat baik. Namun, ada beberapa area yang perlu ditingkatkan, terutama pada precision dan F1-Score.

1) Kelebihan

❖ Akurasi Tinggi

Nilai akurasi sebesar 83.62% menunjukkan bahwa model dapat mengklasifikasikan data dengan benar dalam sebagian besar kasus.

❖ Recall Tinggi

Nilai recall sebesar 91.53% menunjukkan bahwa model sangat efektif dalam mendeteksi kasus positif (atau kondisi penyakit AIDS), yang sangat penting dalam konteks medis untuk memastikan pasien yang benar-benar sakit dapat terdeteksi.

❖ Kesederhanaan dan Interpretabilitas

Algoritma Decision Tree mudah dipahami dan diinterpretasikan, yang memudahkan dalam menjelaskan hasil kepada pihak non-teknis.

2) Kekurangan

❖ Precision Rendah

Nilai precision yang relatif rendah (74.45%) menunjukkan bahwa terdapat cukup banyak false positives. Artinya, model sering mengidentifikasi pasien sebagai positif AIDS padahal sebenarnya tidak.

❖ F1-Score Rendah

Nilai F1-Score yang hanya 68.46% menunjukkan bahwa keseimbangan antara precision dan recall masih kurang optimal. Ini menunjukkan bahwa model masih memiliki ruang untuk perbaikan dalam hal keseimbangan performa.

- ❖ MSE dan RMSE

Nilai MSE dan RMSE menunjukkan bahwa ada variabilitas dalam prediksi model yang bisa ditingkatkan untuk mencapai hasil yang lebih konsisten.

3) Perbaikan yang Dapat Dilakukan di Masa Depan

- ❖ Meningkatkan Preciso

Salah satu fokus utama adalah meningkatkan nilai precision. Ini dapat dicapai dengan berbagai cara, seperti tuning hyperparameter model, menggunakan algoritma yang lebih canggih atau kompleks, atau dengan teknik ensemble (misalnya, Random Forest atau Gradient Boosting).

- ❖ Mengoptimalkan F1-Score

Perbaikan pada precision akan berdampak positif pada F1-Score. Selain itu, dapat dicoba menggunakan metode undersampling atau oversampling untuk mengatasi ketidakseimbangan kelas dalam dataset.

- ❖ Cross-Validation dan Hyperparameter Tuning

Menggunakan teknik cross-validation yang lebih ekstensif dan tuning hyperparameter untuk menemukan konfigurasi model yang optimal.

- ❖ Eksplorasi Algoritma Lain

Meskipun Decision Tree sudah memberikan hasil yang baik, eksplorasi algoritma lain seperti Random Forest, Gradient Boosting Machines (GBM), atau Support Vector Machines (SVM) mungkin akan memberikan performa yang lebih baik.

- ❖ Feature Engineering

Meningkatkan kualitas data dengan melakukan feature engineering, seperti menciptakan fitur baru atau melakukan seleksi fitur untuk mempertahankan fitur yang paling informatif.

- ❖ Penanganan Outliers dan Missing Values

Memastikan bahwa data sudah dibersihkan dengan baik dari outliers dan missing values sebelum digunakan untuk pelatihan model.

Dengan melakukan perbaikan di atas, diharapkan performa model Decision Tree atau model lainnya dalam mengklasifikasikan data penyakit AIDS dapat meningkat di masa yang akan datang, memberikan hasil yang lebih akurat dan andal.

DAFTAR PUSTAKA

- [1] s. d. a. S. M, "TALK SHOW TENTANG PENCEGAHAN HIV/ AIDS PADA REMAJA DI DESA BIALO KECAMATAN GANTARANG KABUPATEN BULUKUMBA," *Jurnal ABDIMAS Panrita*, pp. 1-4, 2020.
- [2] d. Pittara, "Pengertian HIV dan AIDS," *Alodokter*, 2024 .
- [3] P. S. BADUI, "PENULARAN PENYAKIT HIV/AIDS MELALUI TRANSFUSI DARAH," *OSFpreprints*, 2020.
- [4] S. A. Salloum, "Mining in Educational Data:," *Mining in Educational Data: Review and Future Directions*, p. 93, 2020.
- [5] R. Wolff, "5 Types of Classification Algorithms in Machine Learning," *Monkeylearn.Blog*, 2020.
- [6] M. N. & K. D. Yatimah, "Implementasi Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa STIMIK ESQ Menggunakan Decision Tree C4. 5," *In SNIA (Seminar Nasional Informatika dan Aplikasinya)*, pp. Vol. 5, pp. D21-25, 2021, October.
- [7] A. Rakhman, "Prediksi Ketepatan Kelulusan Mahasiswa menggunakan Metode Decision Tree Berbas Particle Swarm Optimization (PSO).," *Smart Comp: Jurnalnya Orang Pintar Komputer*, pp. 6(1), 193-19 2017.
- [8] E. P. Rohmawan, "Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Desicion Tree Dan Artificial Neural Network," *J. Ilm. Matrik*, pp. 20(1), 21-30., 2018.
- [9] A. H. M. H. & S. P. Prasetyo, "Simulasi Penerapan Metode Decision Tree (C4. 5) Pada Penentuan Status Gizi Balita," *Jurnal Nasional Komputasi Dan Teknologi Informasi*, p. 4(3), 2021.
- [10] E. B. & N. A. Santoso, "Analisis sentimen calon presiden indonesia 2019 berdasarkan komentar pub di facebook," *Jurnal Eksplora Informatika*, pp. 9(1), 60-69, 2019.
- [11] N. & F. N. Suryadinata, "Analisis proses berpikir anak berkebutuhan khusus (ABK) dalam menyelesaikan masalah matematika di SMP Inklusi Kota Metro (Studi kasus pada siswa tunagrahita ringan).," *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, pp. 5(1), 94-104., 2016.
- [12] A. P. I. E. S. W. & O. H. Windarto, "Penerapan Data Mining Menggunakan Algoritma C4. 5 Dalam Mengukur Tingkat Kepuasan Pasien BPJS," *In Prosiding Seminar Nasional Riset Information Science (SENARIS)* , pp. (Vol. 2, pp. 376-385)., 2020, July.
- [13] F. K. S. M. A. L. A. T. M. & M. P. D. Marisa, "Data mining konsep dan penerapannya. Deepublish," 2021.
- [14] M. S. R. M. R. & T. A. P. Mustafa, "Implementasi data mining untuk evaluasi kinerja akademik mahasiswa menggunakan algoritma naive bayes classifier," *Creative Information Technology Journa* pp. 4(2), 151-162, 2018.
- [15] A. PUTRA, "PROSES DATA TESTING DAN TRAINING PADA MACHINE LEARNING," *Machine Learning*, 2019.
- [16] I. & F. S. Nawangsih, "Prediksi Pengangkatan Karyawan Dengan Metode Algoritma C5. 0 (Studi Kasu Pt. Mataram Cakra Buana Agung)," *Pelita Teknologi*, pp. 16(2), 24-33., 2021.