

PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS *SYNHETIC MINORITY* *OVER-SAMPLING TECHNIQUE (SMOTE)*

Asep Saepul Anwar¹, Ade Irma Purnama Sari², Agus Bahtiar³, Edi Tohidi⁴

^{1,2}Jurusan Teknik Informatika, STMIK IKMI Cirebon

³Jurusan Sistem Informasi, STMIK IKMI Cirebon

⁴Jurusan Komputerisasi Akuntansi, STMIK IKMI Cirebon

e-mail: ¹azp.jago@gmail.com, ²irma2974@yahoo.com,

³agusbahtiar038@gmail.com, ⁴editohidi00@gmail.com

Intisari

Stunting merupakan tanda dari masalah gizi buruk yang serius dan dapat mengakibatkan balita memiliki tinggi badan pendek dan mempengaruhi pertumbuhan dan perkembangan balita. Data dalam penelitian menggunakan data antropometri balita dari puskesmas Majasem di bulan September 2024, total ada 1368 data balita yang telah direkam pada rentang waktu tersebut. Model *Decision Tree* yang digunakan bertujuan untuk memprediksi status balita, berdasarkan kategori *stunting*. Namun, algoritma *Decision Tree* seringkali menghadapi masalah akurasi yang tidak optimal dikarenakan oleh ketidakseimbangan data kelas dalam dataset. Metode *SMOTE* digunakan sebagai upaya untuk mengatasi masalah ketidakseimbangan kelas, dengan begitu kelas di dalam dataset akan seimbang. Penelitian ini berhasil membuktikan bahwa metode *SMOTE* mampu meningkatkan akurasi model *Decision Tree*, akurasi model dengan *SMOTE* terbaik adalah sebesar 98,56% pada *data training* dan 96,91% pada *data testing* dengan proporsi data 80:20. Penelitian ini berguna untuk membantu tenaga kesehatan untuk memberikan wawasan yang lebih baik terhadap status gizi balita, selain itu dengan model yang dikembangkan bisa meningkatkan efektivitas intervensi kesehatan masyarakat.

Kata kunci—3-5 kata kunci, Abstrak, Template, Jurnal Dinamika Informatika

**PREDIKSI STUNTING MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
(Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

Abstract

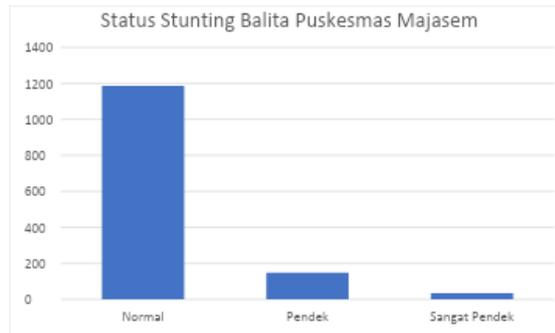
Stunting is a sign of a serious malnutrition problem and can cause toddlers to have a short height and affect the growth and development of toddlers. The data in the study used anthropometric data of toddlers from the Majasem Health Center in September 2024, a total of 1368 toddler data that had been recorded in that time span. The Decision Tree model used aims to predict the status of toddlers, based on the stunting category. However, the Decision Tree algorithm often faces the problem of non-optimal accuracy due to the imbalance of class data in the dataset. The SMOTE method is used as an effort to overcome the problem of class imbalance, so that the classes in the dataset will be balanced. This research successfully proves that the SMOTE method is able to improve the accuracy of the Decision Tree model, the accuracy of the model with the best SMOTE is 98.56% on training data and 96.91% on testing data with a data proportion of 80:20. This research is useful for helping health workers to provide better insight into the nutritional status of toddlers, besides that the developed model can increase the effectiveness of public health interventions.

Keywords— *Stunting, SMOTE, Decision Tree*

PENDAHULUAN

Kontribusi keilmuan Informatika terhadap berbagai aspek kehidupan manusia telah berkembang dengan sangat pesat, termasuk bidang kesehatan. Manfaat dari perkembangan Informatika bagi dunia kesehatan membantu dalam upaya pelayanan kesehatan yang semakin dikembangkan dengan bantuan teknologi *artificial intelligence, machine learning*, dan *Data mining* dapat membantu tenaga kesehatan untuk menangani penyakit pasien. Proses analisis kesehatan yang dilakukan oleh dokter setiap mengobati pasien, membuat volume data kesehatan semakin besar, metode *machine learning* menjadi metode yang cocok untuk mengolah data kesehatan tersebut. Data kesehatan yang diolah bisa digunakan untuk meningkatkan proses diagnosis dan pengobatan penyakit yang lebih baik lagi,

dengan menggunakan metode algoritma *Decision Tree* dapat membantu untuk pengambilan keputusan [1], [2], [3]. Masalah *stunting* balita merupakan masalah kesehatan mendesak yang memerlukan perhatian serius, penelitian ini berfokus kepada implementasi *Decision Tree* yang dioptimasi oleh *SMOTE* untuk menangani masalah ini.



Gambar 1. Sebaran status *stunting* balita di Puskesmas Majasem

Model yang dibuat sering menjadi bias dan tidak akurat disebabkan oleh ketidakseimbangan kelas di dalam data *stunting*. Model yang bias biasanya akan cenderung memprediksi kelas mayoritas dan mengacuhkan kelas minoritas sebagaimana divisualisasikan pada gambar 1. Masalah ini dapat menyebabkan model memiliki akurasi prediksi yang buruk. Penelitian sebelumnya mengungkapkan bahwa untuk menangani masalah ini dapat menggunakan metode *SMOTE*. Metode *SMOTE* bekerja dengan cara menghasilkan contoh sintesis dari kelas minoritas, sehingga representasi data minoritas dan akurasi model dapat ditingkatkan [4], [5]. Kesenjangan literatur mengenai penerapan *SMOTE* dalam konteks kesehatan anak, khususnya dalam analisis *stunting*. Dengan meningkatnya prevalensi *stunting* di tiap negara di dunia khususnya di Indonesia metode yang efektif sangat penting untuk dikembangkan sebagai alat untuk memprediksi dan menganalisis kondisi ini. Target penelitian ini adalah mengisi kesenjangan literatur yang mempraktikkan teknik *SMOTE* dengan *Decision Tree* dan mengevaluasi dampaknya terhadap akurasi prediksi *stunting*. Dengan begitu, penelitian ini tidak hanya berguna bagi dunia akademik, tetapi memiliki peran dalam membantu dalam intervensi yang tepat untuk penanganan *stunting*.

Pendekatan dalam analisis *stunting* telah dilakukan di beberapa penelitian sebelumnya. Penelitian sebelumnya menghadapi masalah ketidakseimbangan data

**PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
(Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

dalam penerapan algoritma *Decision Tree* untuk memprediksi *stunting* [6], [7]. Penelitian oleh Bitew et al. (2021) Menggunakan berbagai metode *machine learning* dalam penelitian mereka untuk memprediksi malnutrisi pada anak. Dataset yang digunakan sebanyak 9471 anak yang diprediksi. Dari 5 model algoritma yang digunakan, *xgbtree* menunjukkan beragam prediktor penting dari kekurangan gizi di tiga hasil yang meliputi waktu menuju sumber air, riwayat anemia, usia anak di atas 30 bulan, ukuran bayi yang kecil dan berat badan ibu yang kurang, dan lain-lain. Algoritma *xgbTree* adalah algoritma terbaik untuk memprediksi kekurangan gizi di Ethiopia dibandingkan dengan algoritma *machine learning* lainnya yang dipertimbangkan dalam penelitian ini. Temuan ini mendukung peningkatan akses terhadap pasokan air, ketahanan pangan, dan regulasi kesuburan, antara lain, dalam upaya meningkatkan gizi anak di Ethiopia [8]. Selain itu penelitian oleh ndagijima et al. (2023) tertarik meneliti masalah *stunting* di negara Rwanda yang memiliki prevalensi *stunting* sebesar 38% di tahun 2015. Penelitian oleh ndagijima et al. bertujuan untuk mengembangkan model untuk memprediksi *stunting* anak-anak di Rwanda. Model terbaik dikembangkan dengan algoritma pengklasifikasi *gradient boosting*, dengan akurasi pelatihan sebesar 80.49% berdasarkan indikator kinerja beberapa model. Berdasarkan *Confusion matrix* , akurasi tes, sensitivitas, spesifisitas, dan F1 dihitung, menghasilkan kemampuan model untuk mengklasifikasikan kasus *stunting* dengan benar sebesar 79.33%, mengidentifikasi anak *stunting* dengan tepat sebesar 72,51%, dan mengkategorikan anak *non-stunting* dengan tepat sebesar 94.49%, dengan luas area di bawah kurva 0,89. Model ini menemukan bahwa tinggi badan ibu, televisi, usia anak, provinsi, pendidikan ibu, berat badan lahir, dan ukuran bayi yang dilahirkan merupakan prediktor terpenting status *stunting* [4]. Penelitian lainnya oleh Yunus et al. menerapkan algoritma *Decision Tree c4.5* untuk memprediksi *stunting* anak usia 0-60 bulan berdasarkan perhitungan *z-score* dengan menggunakan data sebanyak 224 data balita. Hasil dari penelitian tersebut didapatkan pohon keputusan *C4.5* dimana variabel Usia berpengaruh terhadap klasifikasi *stunting* dengan *Gain ratio* tertinggi

yaitu 0.185016337. Sementara itu, evaluasi model menggunakan *Confusion matrix* menghasilkan akurasi sebesar 61.82% [9].

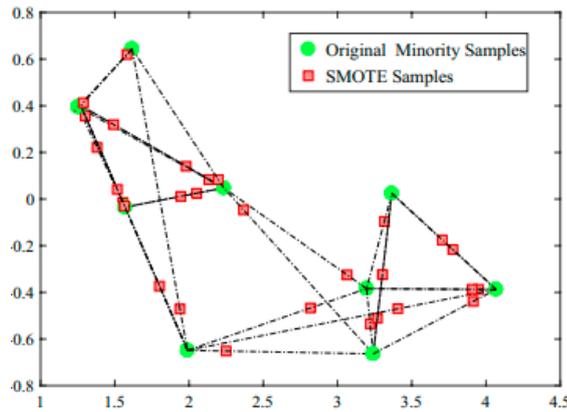
Penelitian ini bertujuan untuk mengevaluasi efektivitas *SMOTE* dalam meningkatkan akurasi model prediksi *stunting* menggunakan algoritma *Decision Tree*. Hasil penelitian memiliki implikasi penting bagi dunia kesehatan terutama pemahaman tentang faktor-faktor *stunting*. Temuan dari penelitian ini bisa menjadi acuan tenaga kesehatan dalam merumuskan intervensi *stunting* yang tepat, serta menjadi dasar yang kuat untuk dijadikan kebijakan masyarakat. Selain itu, penelitian ini berkontribusi pada pengembangan teknik pemodelan, terutama penanganan ketidakseimbangan data. Dengan manfaat bagi pengetahuan akan faktor-faktor penyebab *stunting*, diharapkan akan berdampak kepada tingkat kualitas hidup dan kesehatan anak-anak khususnya di Puskesmas Majasem agar lebih baik lagi dan mengurangi prevalensi *stunting* di masyarakat. Penelitian ini menjadi jalan bagi peneliti lainnya yang tertarik dengan metode penelitian ini dalam analisis kesehatan masyarakat [10], [11].

METODE PENELITIAN

Metode yang digunakan mengadopsi penerapan *SMOTE* untuk menangani ketidakseimbangan dataset *stunting* balita, lalu menerapkan algoritma *Decision Tree* untuk memprediksi status *stunting* balita. Proses penelitian dimulai dengan pengumpulan data antropometri balita di Puskesmas Majasem. Langkah pertama dalam analisis data adalah tahap *data preprocessing* untuk membersihkan dan menyiapkan data untuk dianalisis. Selanjutnya, metode *SMOTE* akan diterapkan untuk menghasilkan contoh sintesis dari kelas minoritas, dengan metode tersebut dapat menambah representasi data [12]. Setelah itu, algoritma *Decision Tree* diterapkan untuk membangun model prediksi, lalu dievaluasi menggunakan akurasi, *recall*, dan *precision*. Pendekatan ini diharapkan dapat menjadi wawasan penting mengenai faktor-faktor penyebab *stunting* dan meningkatkan akurasi prediksi model [13], [14], [15].

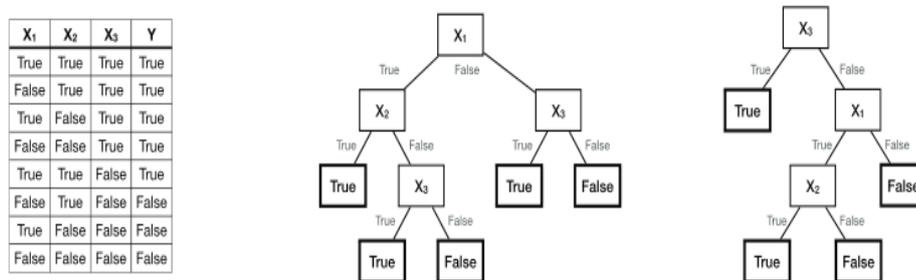
PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)*
 (Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)

Metode SMOTE, bekerja dengan cara menghasilkan sampel sintetis dari kelas minoritas melalui interpolasi linear. Prosesnya diawali dengan pemilihan satu sampel secara acak dari kelas minoritas untuk sebagai acuan dalam mencari tetangga terdekatnya. Teknik *SMOTE* mencari sampel baru yang diciptakan dengan cara mengambil titik antara sampel yang telah dipilih [12]. Mekanisme metode *SMOTE* ditampilkan di gambar 3.5 dibawah ini.

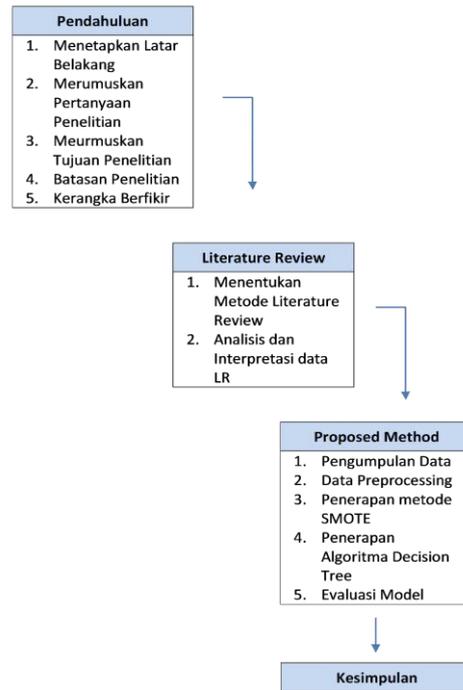


Gambar 2. Mekanisme interpolasi *SMOTE*

Algoritma *Decision Tree* bekerja dengan prinsip basis pembagian (*partitioning*). *Decision Tree* merepresentasikan data dalam bentuk pohon, setiap *node* mewakili suatu fitur atau atribut, setiap cabang menyubstitusi atribut buatan tersebut, dan setiap *leaf node* (*node* daun) mewakili label kelas atau nilai *output*. Untuk mengukur kriteria pemilihan berdasarkan pengukuran informasi seperti *information gain*, *gini index* dan *gain ratio* [16].



Gambar 3. Cara kerja *Decision Tree*



Gambar 4. Metode Penelitian

Metode penelitian yang ditampilkan gambar 4 pada tahap *proposed method* dimulai dengan pengumpulan data sampai evaluasi model. Berikut penjabaran terkait *proposed method*.

Tabel 1. Penjelasan Metode Penelitian

Tahapan	Aktivitas	Deskripsi Aktivitas
1. <i>Proposed Method</i>	Pengumpulan data	Data yang digunakan dikumpulkan dari tiap Posyandu di Puskesmas Majasem di bulan September 2024, total ada 1368 balita pada rentang waktu tersebut yang berhasil direkam.
	<i>Data Preprocessing</i>	Tahap ini menangani masalah di dalam data seperti duplikat, nilai yang hilang, <i>feature scaling</i> , dan <i>encoding</i> .

**PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
 (Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

Penerapan metode <i>SMOTE</i>	Pada atribut target kelas Normal menjadi kelas mayoritas, sementara kelas Pendek dan Sangat Pendek menjadi kelas minoritas, perlu menerapkan teknik <i>over-sampling</i> menggunakan metode <i>SMOTE</i> .
Penerapan Algoritma <i>Decision Tree</i>	Setelah tahapan sebelumnya, algoritma <i>Decision Tree</i> diterapkan untuk menciptakan model yang bisa memprediksi status <i>stunting</i> balita.
Evaluasi Model	Model <i>Decision Tree</i> yang sudah dilatih, perlu untuk dievaluasi untuk akurasi, <i>recall</i> , <i>precision</i> , dan <i>F1-score</i> .

HASIL DAN PEMBAHASAN

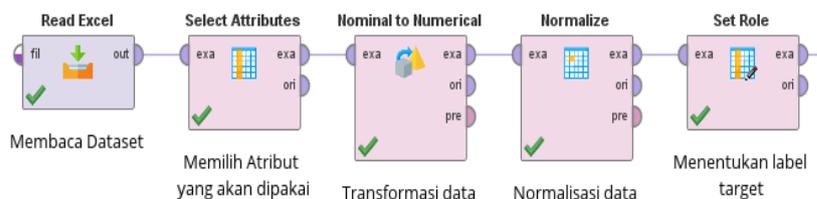
Data diperoleh dari 23 Posyandu di Puskesmas Majasem yang mengumpulkan data melalui penimbangan balita. Data ini berisi indeks antropometri balita dengan total 1368 data, dan bertambah menjadi 3555 baris data setelah melalui teknik *SMOTE*. Tabel 2 berikut menunjukkan atribut dataset

Tabel 2. Atribut dataset

No	Nama atribut	Tipe data
1	No	Integer
2	NIK	Polynomial
3	Nama	Polynomial
4	jk	polynomial
5	Tgl_lahir	Date
6	Nama_ortu	Polynomial
7	Prov	Polynomial
8	Kab/kota	Polynomial
9	Kec	Polynomial
10	Puskesmas	Polynomial

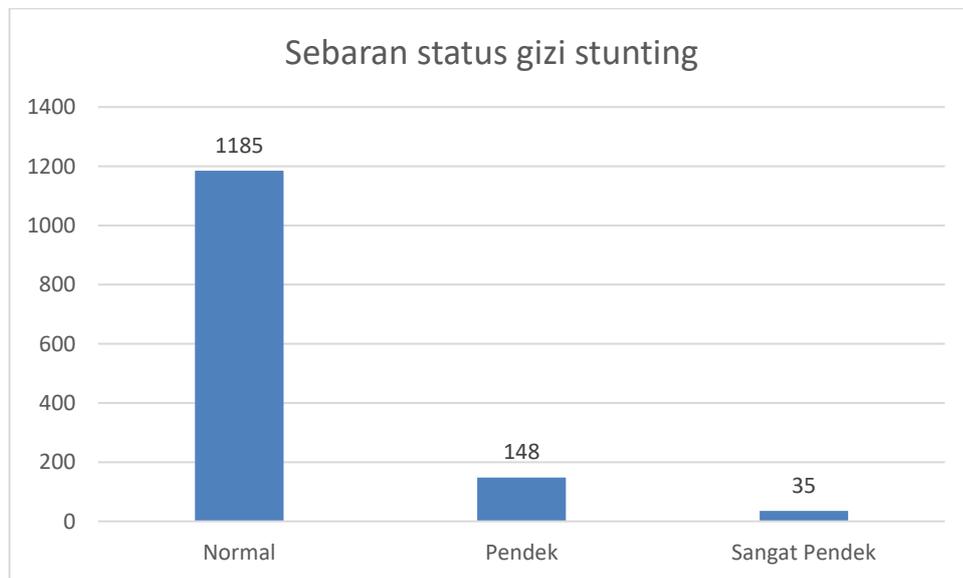
11	Desa/kel	Polynomial
12	Posyandu	Polynomial
13	Rt	Integer
14	Rw	Integer
15	Alamat	Polynomial
16	Tgl_pengukuran	Date
17	Bb	Real
18	Tb	Real
19	Cara ukur	Polynomial
20	Bb/u	Polynomial
21	Zs bb/u	Real
22	Tb/u	Polynomial
23	Zs tb/u	Real
24	Bb/tb	Polynomial
25	Zs bb/tb	Real
26	naik_berat_badan	Polynomial
27	Usia_saat_ukur	Integer

Proses analisis data menggunakan perangkat lunak *RapidMiner*. Tahap awal dalam proses analisis adalah melakukan *preprocessing* untuk membersihkan data balita agar siap dilanjutkan ke tahap pemodelan. Teknik *data Preprocessing* yang akan dilakukan seperti *data cleansing*, *data normalization*, dan *data transformation* [17]. Operator yang digunakan ditunjukkan pada gambar 5 berikut.



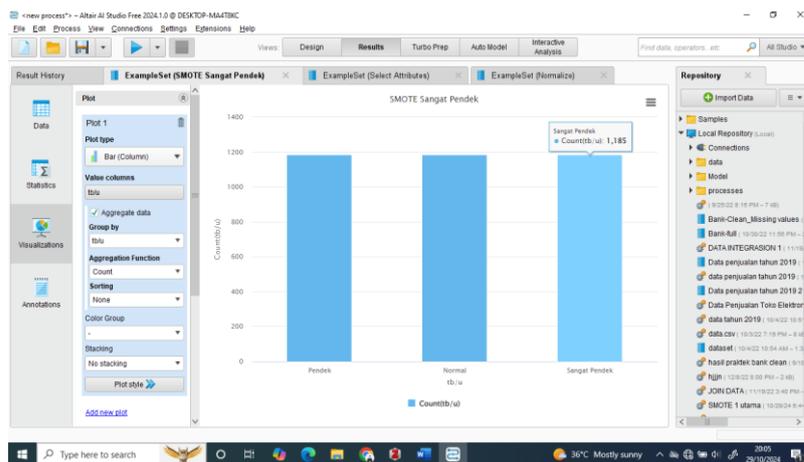
Gambar 5. Proses *data preprocessing*

PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS *SYNHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)*
(Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)



Gambar 6. Sebaran status gizi *stunting*

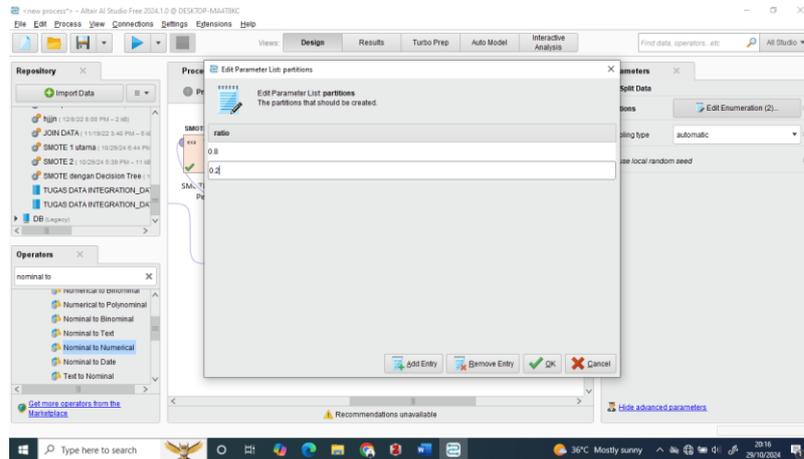
Pada gambar 6, menunjukkan bahwa kelas normal merupakan kelas mayoritas dengan total sebanyak 1185 data, kelas pendek sebanyak 148 data, dan kelas sangat pendek sebanyak 35 data. Teknik *SMOTE* dilakukan pada kelas minoritas adalah kelas pendek dan sangat pendek, hal ini akan memperbanyak jumlah kedua kelas tersebut sama dengan kelas normal. Hasil teknik *SMOTE* ditampilkan pada gambar 7 dibawah ini.



Gambar 7. Hasil teknik *SMOTE*

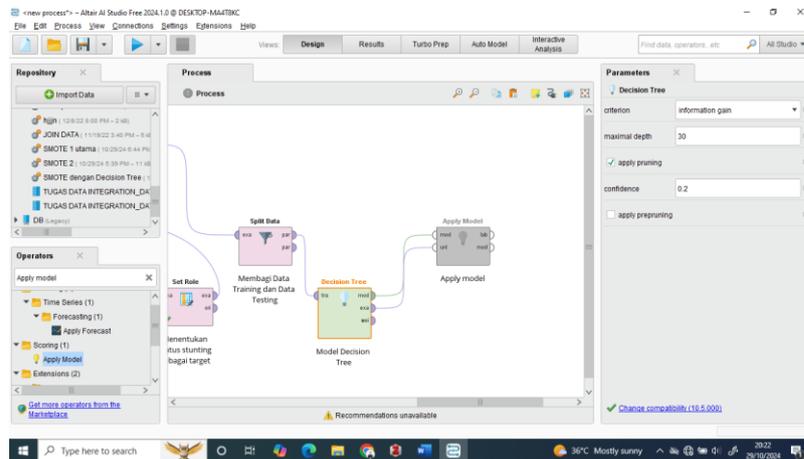
Implementasi algoritma *Decision Tree* membutuhkan beberapa operator. Operator pertama yang dibutuhkan adalah *split Data*, untuk membagi data berdasarkan proporsi. Proporsi pembagian dataset dengan metode *split data* yang

akan dicoba adalah 50:50, 60:40, 70:30, 80:20 dan 90:10. Proses *split data* dapat diamati melalui gambar 8 berikut.



Gambar 8. Pembagian dataset untuk *training* dan *testing*

Proses selanjutnya adalah menerapkan algoritma *Decision Tree* menggunakan operator yang bernama *Decision Tree*. Selain menggunakan operator *Decision Tree*, juga membutuhkan operator *Apply Model* untuk mengimplementasikan model pada *data testing*. Perhatikan gambar 8 di bawah ini, yang menjelaskan penerapan algoritma *Decision Tree*.



Gambar 9. Implementasi *Decision Tree*

Operator yang digunakan pada tahap evaluasi model *Decision Tree* adalah *Performance(Classification)*, operator ini berfungsi untuk mengevaluasi performa model klasifikasi. Hasil pemodelan dapat dilihat pada pembahasan berikut.

A. Penerapan algoritma *Decision Tree* untuk memprediksi data *stunting* balita

**PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
(Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

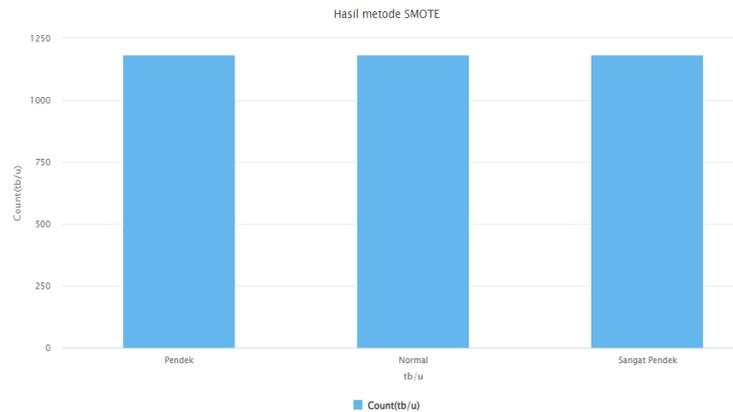
Metode algoritma *Decision Tree* untuk mengklasifikasikan status *stunting* balita dikombinasikan dengan teknik *SMOTE*. Proses pemodelan ini menggunakan proporsi pembagian dataset 50:50, 60:40, 70:30, 80:20, dan 90:10. Tabel 3 berikut mengungkapkan evaluasi model.

Tabel 3. Hasil akurasi model *Decision Tree*

Proporsi	Decision Tree tanpa <i>SMOTE</i>		Decision Tree dengan <i>SMOTE</i>	
	<i>Training</i>	<i>Testing</i>	<i>Training</i>	<i>Testing</i>
50:50	95,18%	93,57%	97,58%	98,31%
60:40	93,54%	87,93%	88,09%	98,17%
70:30	92,39%	88,26%	88,11%	97,28%
80:20	94,06%	90,15%	98,56%	96,91%
90:10	93,75%	88,76%	98,50%	94,63%

Pada tabel 3 terlihat bahwa 2 kategori yang digunakan untuk memprediksi data status *stunting* balita di Puskesmas Majasem. Untuk kategori model dengan *SMOTE* akurasi terbaik terletak pada proporsi dataset 80:20 dibuktikan dengan akurasi *data training* dengan nilai 98.56% dan akurasi *data testing* dengan nilai 96.91%. Sedangkan kategori model tanpa *SMOTE* akurasi terbaik pada proporsi 50:50 dibuktikan dengan akurasi *data training* dengan nilai 95.18% dan pada *data testing* dengan nilai 93.57%.

B. Menerapkan *SMOTE* untuk meningkatkan performa model memprediksi *stunting* pada balita di Puskesmas Majasem



Gambar 10. Hasil *SMOTE*

Pada gambar 10, setelah melakukan metode *over-sampling* menggunakan teknik *SMOTE* membuat kelas dalam dataset menjadi seimbang. Kelas Pendek dan Sangat Pendek masing-masing kelas menjadi memiliki 1185 data mengikuti jumlah kelas mayoritas Normal yang memiliki 1185 baris data. Sehingga total data menjadi 3555 baris data.

Tabel 4. Peningkatan akurasi setelah menerapkan metode *SMOTE*

Proporsi	Decision Tree tanpa <i>SMOTE</i>		Decision Tree dengan <i>SMOTE</i>		Peningkatan Akurasi setelah metode <i>SMOTE</i>	
	Training	Testing	Training	Testing	Training	Testing
50:50	95,18%	93,57%	97,58%	98,31%	+ 2,40%	+ 4,74%
60:40	93,54%	87,93%	88,09%	98,17%	- 5,45%	+ 10,24%
70:30	92,39%	88,26%	88,11%	97,28%	- 4,28%	+ 9,02%
80:20	94,06%	90,15%	98,56%	96,91%	+ 4,50%	+ 6,76%
90:10	93,75%	88,76%	98,50%	94,63%	+ 4,75%	+ 5,87%

Pada tabel 4 menunjukkan bahwa metode *SMOTE* secara konsisten meningkatkan akurasi pada *data testing* di semua proporsi pembagian data, dengan peningkatan tertinggi sebesar 10.24% pada proporsi 60:40. Namun, pada *data training* memberikan peningkatan yang bervariasi, peningkatan terjadi pada proporsi 50:50, 80:20, dan 90:10, namun terjadi penurunan akurasi pada proporsi

**PREDIKSI *STUNTING* MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
 (Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

60:40 serta 70:30. Oleh karena itu, metode *SMOTE* efektif meningkatkan akurasi pada *data testing*, hal ini membuktikan bahwa peran *SMOTE* penting mengatasi masalah data yang tidak seimbang agar model dapat lebih general terhadap data baru.

C. Membandingkan akurasi prediksi antara model yang menggunakan *SMOTE* dan model tanpa teknik *SMOTE*

Tabel 4 menunjukkan kategori akurasi *data training* model dengan teknik *SMOTE* dan model tanpa teknik *SMOTE*. Dari tabel tersebut diketahui bahwa akurasi model terbaik tanpa menggunakan metode *SMOTE* adalah pada proporsi pembagian dataset 50:50, sedangkan pada akurasi model yang menggunakan metode *SMOTE* adalah model dengan proporsi pembagian data 80:20. Penjelasan lebih detail mengenai *Confusion matrix* model terbaik pada 2 kategori berbeda tersebut, dijelaskan di bawah ini.

a. Performa Model tanpa *SMOTE* proporsi 50:50

accuracy: 95.18%				
	true Pendek	true Normal	true Sangat Pendek	class precision
pred. Pendek	70	19	8	72.16%
pred. Normal	4	573	1	99.13%
pred. Sangat Pendek	0	1	8	88.89%
class recall	94.59%	96.63%	47.06%	

Gambar 11. *Confusion Matrix* model tanpa *SMOTE* data training 50%

Pada gambar 11 menampilkan *Confusion matrix* model tanpa *SMOTE* dengan *data training* sebesar 50% atau 684 baris data. Terlihat bahwa kelas Normal memiliki data yang lebih banyak dari pada lainnya, karena ini merupakan kelas mayoritas, dan belum menerapkan metode *SMOTE* sehingga data belum seimbang. *Recall* tertinggi adalah pada kelas normal dengan akurasi 96,63% namun diprediksi sebagai pendek sebanyak 19 data dan sebagai sangat pendek sebanyak 1 kali, kelas Pendek memiliki akurasi *recall* sebesar 94.59% diprediksi salah sebagai normal sebanyak normal sebanyak 4 data, dan kelas Sangat Pendek menjadi kelas dengan akurasi *recall* terburuk sebesar 47,06% diprediksi salah sebagai kelas pendek sebanyak 8 data, dan 1 kali sebagai kelas normal, hal ini jelas karena Normal memiliki jumlah kelas yang lebih banyak jika dibandingkan kelas Sangat Pendek yang jauh lebih sedikit. Pada *precision* prediksi Normal memiliki akurasi *precision*

yang sempurna 99,13%, prediksi Sangat pendek dengan akurasi *precision* sebesar 88.89%, dan prediksi Pendek memiliki akurasi *precision* terburuk dengan akurasi hanya 77,16%, hal ini dapat disebabkan karena kelas Sangat Pendek memiliki kelas yang sedikit sehingga kelas Sangat Pendek memiliki jumlah variasi data yang sedikit.

accuracy: 93.57%				
	true Pendek	true Normal	true Sangat Pendek	class precision
pred. Pendek	39	4	4	82.98%
pred. Normal	32	587	0	94.83%
pred. Sangat Pendek	3	1	14	77.78%
class recall	52.70%	99.16%	77.78%	

Gambar 12. *Confusion Matrix* model tanpa *SMOTE* data testing 50%

Pada gambar 12 menampilkan *Confusion matrix* model tanpa *SMOTE* pada *data testing* sebesar 50% atau sebanyak 684 data. Kelas Normal masih menjadi kelas yang memiliki akurasi *recall* yang sempurna yakni 99,16% diprediksi sebagai kelas pendek sebanyak 4 data, dan prediksi sangat pendek sebanyak 1 data, kelas Sangat Pendek dengan nilai *recall* sebesar 77.78% kesalahan prediksi sebagai kelas pendek sebanyak 4 data, sedangkan kelas Pendek menjadi kelas yang memiliki akurasi *recall* terburuk yakni sebesar 52,70% prediksi salah banyak diprediksi sebagai kelas normal sebanyak 32 data dan sebagai kelas sangat pendek sebanyak 3 data. Hal ini dapat disebabkan karena kelas mayoritas, sehingga pada tahap pelatihan model mampu mempelajari data kelas Normal sehingga akurasi *recall* membaik pada *data testing* untuk kelas Normal. Pada *precision* prediksi Normal memiliki akurasi *precision* terbaik sebesar 94,83%, prediksi pendek dengan nilai *precision* 88.89%, dan prediksi Sangat Pendek menjadi prediksi yang memiliki akurasi *precision* terburuk menjadi sebesar 77,78%.

b. Performa Model dengan *SMOTE* proporsi 80:20

accuracy: 98.56%				
	true Pendek	true Normal	true Sangat Pendek	class precision
pred. Pendek	928	13	7	97.89%
pred. Normal	15	935	1	98.32%
pred. Sangat Pendek	5	0	940	99.47%
class recall	97.89%	98.63%	99.16%	

Gambar 13. *Confusion Matrix* model dengan *SMOTE* data training 80%

**PREDIKSI STUNTING MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
 (Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

Pada gambar 13 menggunakan proporsi 80:20. Menunjukkan *confusion matrix* untuk model dengan metode *SMOTE* dengan *data training* sebesar 80% atau sekitar 2844 baris data. Terlihat bahwa untuk *recall* pendek sebesar 97,89% merupakan kelas yang lebih banyak salah diprediksi sebagai prediksi normal sebanyak 15 data, dan prediksi sangat pendek sebanyak 5 data. *Recall* sangat pendek merupakan kelas dengan akurasi *recall* terbaik sebesar 99,16%, dimana 7 data diprediksi sebanyak Pendek, dan 1 data diprediksi sebagai normal. Sedangkan kelas Normal dengan nilai *recall* sebesar 96.63% diprediksi sebagai pendek sebanyak 13 data, dan tidak ada diprediksi sebagai sangat pendek.

accuracy: 96.91%				
	true Pendek	true Normal	true Sangat Pendek	class precision
pred. Pendek	231	9	6	93.90%
pred. Normal	2	228	1	98.70%
pred. Sangat Pendek	4	0	230	98.29%
class recall	97.47%	96.20%	97.05%	

Gambar 14. *Confusion Matrix* model dengan *SMOTE* data testing 20%

Pada gambar 4.25, dengan proporsi 80:20. *Data testing* yang digunakan sebesar 20% atau sebanyak 711 baris data, memiliki akurasi total sebesar 96,91%. *Precision* prediksi pendek memprediksi salah sebanyak 9 untuk kelas Normal dan 6 untuk kelas Sangat Pendek, menjadi nilai *precision* terburuk dengan akurasi *precision* sebesar 93,90%. Prediksi normal memprediksi salah sebanyak 2 data sebagai normal, dan 1 data sebagai sangat pendek, memperoleh nilai *precision* sebesar 98.70%. Sementara prediksi sangat pendek memprediksi salah hanya kepada kelas pendek sebanyak 4 data, memperoleh nilai *precision* sebesar 98.29%. Sementara untuk *recall*, nilai kelas Normal menjadi kelas yang paling banyak salah diprediksi sebagai prediksi Pendek sebanyak 9 data, mendapatkan nilai akurasi *recall* sebesar 96,20%.

KESIMPULAN

1. Dengan menerapkan algoritma *Decision Tree* menunjukkan bahwa penggunaan metode *SMOTE* mampu meningkatkan akurasi, kenaikan akurasi paling mencolok terletak pada proporsi *split data* yang lebih tinggi yakni pada 80:20

dan 90:10. Model dengan *SMOTE* mencapai akurasi tertinggi pada proporsi 80:20 dengan nilai akurasi 98,56% pada *data training* dan 96,91% pada *data testing*. Tanpa *SMOTE* akurasi cenderung rendah di semua proporsi. Hal ini membuktikan bahwa metode *SMOTE* mampu meningkatkan generalisasi model *Decision Tree* pada data status *stunting* balita.

2. Dengan menerapkan *SMOTE* pada algoritma *Decision Tree* meningkatkan akurasi model dengan signifikan dengan menangani ketidakseimbangan data secara efektif. Dengan metode *SMOTE*, meningkatkan kenaikan akurasi tertinggi pada *data testing* sebesar 10,24% pada proporsi data 60:40. Sedangkan pada *data training* kenaikan akurasi terbesar 4,72% pada proporsi 90:10.
3. Dari penerapan *Decision Tree* tanpa *SMOTE* dan dengan *SMOTE*, terlihat bahwa model dengan *SMOTE* secara konsisten memberikan performa yang lebih baik pada *data testing* dibandingkan tanpa *SMOTE*. Pada model tanpa *SMOTE* akurasi yang terbaik terletak pada model dengan proporsi 50:50, model ini memiliki akurasi *data training* sebesar 95,18% dan *data testing* sebesar 93,57%. Sementara itu untuk model yang dikombinasikan dengan metode *SMOTE*, memiliki performa yang terbaik berada pada model pada proporsi 80:20 yang memiliki performa yang baik pada *data training* dengan akurasi 98,56%, akurasi pun membaik pada *data testing* dengan akurasi sebesar 96,91%.

SARAN

Implementasi metode *SMOTE* perlu dilakukan eksperimen dengan pengaturan parameter seperti penentuan k-neighbour atau tingkat over-sampling. Hal ini agar metode tersebut dapat menghasilkan data sintesis yang lebih representatif, sehingga dapat mengurangi kesalahan pada kelas-kelas tersebut.

Metode sampling lainnya yang disarankan untuk digunakan adalah Adaptive Synthetic Sampling (ADASYN). Metode ini bekerja dengan menghasilkan sampel dengan bobot yang lebih tinggi, sehingga data sintetik yang dihasilkan lebih efektif untuk kelas minoritas.

**PREDIKSI STUNTING MENGGUNAKAN ALGORITMA *DECISION TREE* BERBASIS
SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE)
(Asep Saepul Anwar, Ade Irma Purnama Sari, Agus Bahtiar, Edi Tohidi)**

Untuk meningkatkan generalisasi model, perlu untuk menggunakan populasi data dari daerah lain atau pada periode waktu di masa depan. Data tersebut memiliki sifat yang berbeda dengan data yang dilatih karena kondisi lingkungan serta variasi populasi dapat mempengaruhi hasil pengukuran antropometri balita.

UCAPAN TERIMA KASIH

Ucapan terimakasih penulis sampaikan kepada rektor kampus STMIK IKMI Cirebon yang telah meringankan beban finansial bagi penulis. Tidak lupa penulis ucapkan terima kasih banyak kepada dosen pembimbing Ibu Ade Irma Purnama Sari, M.Kom dan Bapak Agus Bahtiar, M.Kom yang dengan sabar membimbing saya sampai penelitian ini selesai.

DAFTAR PUSTAKA

- [1] J. Kim, S. Mun, S. Lee, K. Jeong, and Y. Baek, "Prediction of metabolic and pre-metabolic syndromes using machine learning models with anthropometric, lifestyle, and biochemical factors from a middle-aged population in Korea," *BMC Public Health*, vol. 22, no. 1, p. 664, Apr. 2022, doi: 10.1186/s12889-022-13131-x.
- [2] Z.-Y. Shi, J.-S. Hon, C.-Y. Cheng, H.-T. Chiang, and H.-M. Huang, "Applying Machine Learning Techniques to the Audit of Antimicrobial Prophylaxis," *Applied Sciences*, vol. 12, no. 5, p. 2586, Mar. 2022, doi: 10.3390/app12052586.
- [3] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmaddeni, and D. Wulandari, "Implementasi Algoritma Decision Tree untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix: Implementation of Decision Tree Algorithm for Movie Recommendation and Rating Classification on the Netflix Platform," *MALCOM*, vol. 4, no. 2, pp. 570–579, Mar. 2024, doi: 10.57152/malcom.v4i2.1255.
- [4] S. NDAGIJIMANA, J. Ntaganda, E. Masabo, and I. Kabano, "Prediction of Stunting Among Under-5 Children in Rwanda Using Machine Learning Techniques," *Journal of Preventive Medicine and Public Health*, 2023, doi: 10.3961/jpmph.22.388.
- [5] S. Sutarmi, W. Warijan, T. Indrayana, D. P. P. B, and I. Gunawan, "Machine Learning Model For Stunting Prediction," *JHS*, vol. 4, no. 9, pp. 10–23, Sep. 2023, doi: 10.46799/jhs.v4i9.1073.
- [6] H. Shen, "Machine Learning Algorithms for Predicting Stunting Among Under-Five Children in Papua New Guinea," *Children*, 2023, doi: 10.3390/children10101638.

- [7] M. Yunus, M. K. Biddinika, and A. Fadlil, "Classification of Stunting in Children Using the C4.5 Algorithm," *Jurnal Online Informatika*, 2023, doi: 10.15575/join.v8i1.1062.
- [8] F. H. Bitew, C. S. Sparks, and S. H. Nyarko, "Machine learning algorithms for predicting undernutrition among under-five children in Ethiopia," *Public Health Nutr.*, pp. 1–12, Oct. 2021, doi: 10.1017/S1368980021004262.
- [9] M. Yunus, M. K. Biddinika, and A. Fadlil, "Classification of Stunting in Children Using the C4.5 Algorithm," *join*, vol. 8, no. 1, pp. 99–106, Jun. 2023, doi: 10.15575/join.v8i1.1062.
- [10] H. Zhu, X. Zheng, and L. Zhao, "Analysis of Teachers' Cognitive Ability and Teaching Motivation on the Academic Achievement of Students in Higher Education via Employment Data Guidance," *Electronics*, 2023, doi: 10.3390/electronics12030572.
- [11] C. В. Новиков, "Data Science and Big Data Technologies Role in the Digital Economy," *Tem Journal*, 2020, doi: 10.18421/tem92-44.
- [12] D. Elreedy, A. F. Atiya, and F. Kamalov, "A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Mach Learn*, vol. 113, no. 7, pp. 4903–4923, Jul. 2024, doi: 10.1007/s10994-022-06296-4.
- [13] A. Prasetyo, "Simulasi Penerapan Metode Decision Tree (C4.5) Pada Penentuan Status Gizi Balita," *JNKTI*, vol. 4, no. 3, pp. 209–214, Jun. 2021, doi: 10.32672/jnkti.v4i3.2983.
- [14] X. Wang, "Research on Fault Prevention and Maintenance System of Automatic Substation Primary Equipment Based on Decision Tree Algorithm," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, p. 20230235, Jan. 2024, doi: 10.2478/amns.2023.2.00235.
- [15] A. Fauzi and A. H. Yunial, "Optimasi Algoritma Klasifikasi Naive Bayes, Decision Tree, K – Nearest Neighbor, dan Random Forest menggunakan Algoritma Particle Swarm Optimization pada Diabetes Dataset," *JEPIN*, vol. 8, no. 3, p. 470, Dec. 2022, doi: 10.26418/jp.v8i3.56656.
- [16] H. Blockeel, L. Devos, B. Frénay, G. Nanfack, and S. Nijssen, "Decision trees: from efficient prediction to responsible AI," *Front. Artif. Intell.*, vol. 6, p. 1124553, Jul. 2023, doi: 10.3389/frai.2023.1124553.
- [17] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.glt.2022.04.020.