

# Implementation of the Tomek-Link Approach in Machine Learning Models for ABO<sub>3</sub> Perovskite Bandgap Classification

Johana Oktavia Ramadhani<sup>1\*</sup>, Aliyah Zahratu Rizqi<sup>2\*</sup>, Desvita Maharani<sup>3\*</sup>, Muhamad Akrom<sup>4\*\*</sup>

\* Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro

\*\*Research Center for Quantum Computing and Materials Informatics, Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131 Indonesia

[111202315024@mhs.dinus.ac.id](mailto:111202315024@mhs.dinus.ac.id)<sup>1</sup>, [111202315294@mhs.dinus.ac.id](mailto:111202315294@mhs.dinus.ac.id)<sup>2</sup>, [111202315298@mhs.dinus.ac.id](mailto:111202315298@mhs.dinus.ac.id)<sup>3</sup>, [m.akrom@dsn.dinus.ac.id](mailto:m.akrom@dsn.dinus.ac.id)<sup>4</sup>

## Article Info

### Article history:

Received 2026-01-20

Revised 2026-02-27

Accepted 2026-04-10

### Keyword:

Band Gap,  
Class Imbalance,  
Machine Learning,  
Perovskite Oksida,  
Tomek Links.

## ABSTRACT

ABO<sub>3</sub> perovskite oxide materials exhibit significant variations in electronic properties, particularly in their band gap characteristics (direct vs. indirect), which are crucial for optoelectronic applications. Experimental approaches and density functional theory (DFT)-based calculations for determining the band gap type require high costs and computational resources, making Machine Learning (ML) a more efficient alternative. However, the imbalanced class distribution in the perovskite oxide dataset (84% direct and 16% indirect after data cleaning) potentially leads to model bias towards the majority class. This study evaluates the effect of applying Tomek Links, a decision boundary cleaning-based undersampling technique, on the performance of band gap type classification using Multi-Layer Perceptron (MLP), Gradient Boosting, CatBoost, and Extra Trees. The dataset consists of 3,469 samples with six predictor numerical features and a binary classification target. Tomek Links is applied exclusively to the training data in a controlled ML pipeline that includes feature standardization and 5-fold stratified cross-validation. Results show that the application of Tomek Links improves minority class (Indirect) recall by up to +9% and macro F1-score by up to +0.019 compared to the baseline, with minimal changes in global accuracy. Feature importance analysis identifies average ionic character as the primary determinant of classification, consistent with material band structure theory. These findings confirm that Tomek Links is effective as a decision boundary cleaning mechanism to reduce bias towards the majority class and improve model sensitivity in data-driven material exploration.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

## I. PENDAHULUAN

Material oksida perovskite dengan rumus umum ABO<sub>3</sub> merupakan salah satu kelas material kristalin yang banyak diteliti berkat fleksibilitas komposisinya dan variasi sifat elektronik yang bisa dihasilkan [1], [2], [3]. Struktur dasar dari ABO<sub>3</sub> terdiri dari kation A dan B yang menempati lokasi yang berbeda di dalam kisi kristal, di mana atom oksigen mengelilingi kation B dalam suatu konfigurasi octahedral [1], [2]. Variasi unsur di posisi A dan B memberikan kesempatan untuk menghasilkan berbagai karakteristik elektronik, termasuk perbedaan dalam jenis band gap [1], [3], [4], yang membuat material ini relevan untuk berbagai aplikasi seperti perangkat optoelektronik, energi, dan katalisis [2], [4].

Dalam konteks aplikasi tersebut, jenis band gap yang bersifat langsung (direct) maupun tidak langsung (indirect) memegang peranan penting karena berpengaruh terhadap mekanisme transisi elektronik serta efisiensi optik material. Penentuan karakter band gap secara konvensional umumnya dilakukan melalui eksperimen laboratorium maupun perhitungan berbasis *density functional theory* (DFT). Namun, pendekatan-pendekatan ini memerlukan biaya, waktu, dan sumber daya komputasi yang signifikan, terutama ketika diterapkan pada ruang komposisi material yang luas. Kondisi ini menegaskan perlunya pendekatan komputasional yang lebih efisien sebagai alternatif untuk mendukung eksplorasi sifat material [5], [6].

Pendekatan Machine Learning telah banyak dimanfaatkan sebagai solusi berbasis data untuk mempercepat prediksi maupun klasifikasi sifat material [5], [6]. Dengan mempelajari hubungan antara representasi numerik material dan sifat target, ML memungkinkan proses penyaringan kandidat material dilakukan secara cepat dan berskala secara komputasional tanpa memodelkan mekanisme fisika material secara eksplisit [5]. Dalam studi perovskite oksida, Machine Learning telah menunjukkan potensi yang besar untuk membantu dalam klasifikasi jenis band gap sebagai langkah awal dalam penelitian berbasis data mengenai material [6], [7], [8].

Meskipun demikian, penerapan Machine Learning dalam praktik sering dihadapkan pada permasalahan kualitas dan karakteristik data [9]. Salah satu tantangan utama adalah ketidakseimbangan kelas (*class imbalance*) yang umumnya muncul pada distribusi kelas target dan dapat diperparah oleh tumpang tindih distribusi fitur antar kelas di sekitar batas Keputusan [9]. Pada dataset perovskite oksida yang digunakan dalam penelitian ini, distribusi jenis band gap menunjukkan dominasi kelas mayoritas yang cukup menonjol [10]. Ketidakseimbangan semacam ini berpotensi menyebabkan model klasifikasi menjadi bias terhadap kelas mayoritas, sehingga kemampuan model dalam mengenali kelas minoritas kurang terwakili meskipun nilai akurasi keseluruhan tampak tinggi [9].

Berbagai pendekatan telah dikembangkan untuk menangani permasalahan ketidakseimbangan kelas dalam tugas klasifikasi, termasuk teknik oversampling, undersampling, dan pendekatan hibrida, yang masing-masing memiliki implikasi berbeda terhadap distribusi data, kompleksitas model, dan interpretasi hasil [6], [9]. Oleh karena itu, pemilihan dan evaluasi metode penanganan ketidakseimbangan kelas perlu disesuaikan dengan karakteristik permasalahan yang dihadapi agar pipeline klasifikasi yang dibangun tetap andal.

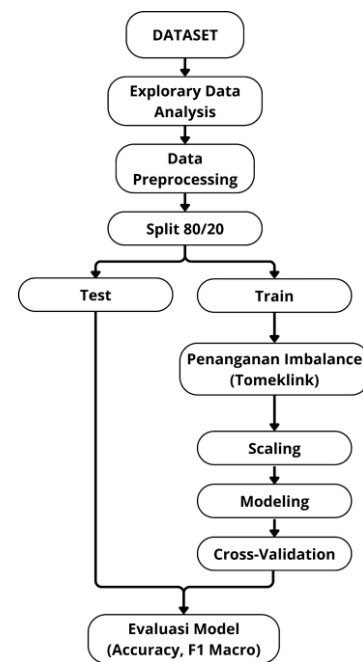
Salah satu teknik undersampling yang banyak digunakan adalah Tomek Links, yang bekerja dengan mengidentifikasi pasangan data lintas kelas yang berada sangat dekat dalam ruang fitur, sehingga sampel yang ambigu di sekitar batas keputusan dapat dihapus untuk mengurangi tumpang tindih antar kelas [11], [12]. Meskipun demikian, pengaruh penerapan Tomek Links terhadap performa klasifikasi jenis band gap pada perovskite oksida, khususnya dalam pipeline Machine Learning yang melibatkan standarisasi fitur, masih memerlukan kajian empiris lebih lanjut.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis pengaruh penerapan Tomek Links terhadap performa klasifikasi jenis band gap pada material perovskite oksida  $ABO_3$  yang mengalami ketidakseimbangan kelas. Evaluasi dilakukan melalui perbandingan performa beberapa algoritma klasifikasi dalam pipeline Machine Learning yang terkontrol, dengan menekankan dampak penerapan Tomek Links terhadap kinerja model pada kondisi data tidak seimbang. Hasil penelitian ini diharapkan dapat memberikan kontribusi

metodologis dalam pengembangan pendekatan klasifikasi berbasis data yang lebih andal untuk permasalahan material informatics dengan karakteristik data tidak seimbang.

## II. METODE

Penelitian ini menerapkan pendekatan komputasional berbasis data dengan memanfaatkan metode supervised Machine Learning untuk melakukan klasifikasi jenis band gap pada material perovskite oksida  $ABO_3$ . Alur kerja penelitian disusun secara sistematis dan dirangkum dalam diagram alir pada Gambar 1. Tahapan penelitian meliputi analisis data awal, pra-pemrosesan data, penanganan ketidakseimbangan kelas, standarisasi fitur, proses pemodelan, serta evaluasi kinerja model klasifikasi.



Gambar 1. Diagram Alir Penelitian

### A. Deskripsi Dataset

Dataset yang digunakan dalam penelitian ini bersumber dari studi yang dilakukan oleh Priyanga dkk. [8] berjudul “*Prediction of nature of band gap of perovskite oxides ( $ABO_3$ ) using a machine learning approach*”. Data tersebut berupa data numerik terstruktur (tabular) yang merepresentasikan karakteristik fisik dan kimia dari material perovskite oksida  $ABO_3$ .

Dalam kerangka supervised Machine Learning, dataset diformulasikan sebagai permasalahan klasifikasi biner dengan variabel target berupa jenis band gap, di mana kelas 0 merepresentasikan indirect band gap dan kelas 1 merepresentasikan direct band gap. Klasifikasi ini memiliki signifikansi ilmiah yang krusial bagi desain material optoelektronik [7], [5], [13]. Material direct band gap sangat diperlukan untuk efisiensi emisi foton pada perangkat seperti LED dan laser, sedangkan indirect band gap memiliki

karakteristik transisi elektron yang berbeda yang relevan untuk aplikasi fungsional lainnya [5], [7], [13], [14].

Variabel prediktor terdiri atas fitur-fitur numerik kontinu yang mencakup jari-jari ionik (RA, RB), elektronegativitas (ENA, ENB), serta fitur turunan seperti Avg Iconic Char [8]. Distribusi kelas dalam dataset ini menunjukkan ketidakseimbangan, di mana kelas indirect merupakan minoritas. Hal ini secara fisik mencerminkan distribusi alami dalam database material, di mana konfigurasi atom tertentu pada struktur  $ABO_3$  lebih cenderung membentuk sifat direct band gap dibandingkan indirect [7], [13]. Contoh sampel data yang digunakan ditunjukkan pada Tabel I.

TABEL I  
CONTOH DATASET PEROVSKITE OKSIDA  $ABO_3$

RA	RB	EN A	ENB	ENR	Avg Iconic Char	Band gap
1,05	1,18	1,25	0,95	-3,306393	0,17924226 8051812	0
1,12	0,27	1,1	2,04	-1,1173571	0,14388656 2739263	0
0,8	0,71	1,66	1,3	-1,9350715	0,14873741 4161745	0
0,83	0,78	1,3	1,83	-2,0316787	0,14175260 8472182	0
1,08	0,65	1,22	1,55	-2,0429285	0,15694293 3195844	0
...	...	...	...	...	...	...

### B. Exploratory Data Analysis (EDA) dan Pra-pemrosesan Data

Tahap awal penelitian mencakup proses exploratory data analysis (EDA) dan pra-pemrosesan data. EDA dilakukan untuk memahami karakteristik statistik dan distribusi fitur numerik, serta mengidentifikasi potensi permasalahan pada data. Tahapan pra-pemrosesan meliputi penanganan nilai hilang (missing values) dan penghapusan data duplikat pada fitur numerik.

Nilai pencilan (outlier) tidak dihilangkan dalam penelitian ini dengan tujuan mempertahankan representasi distribusi alami data material. Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan teknik stratified sampling untuk menjaga proporsi kelas pada kedua subset. Data uji digunakan secara eksklusif pada tahap evaluasi akhir guna memperoleh estimasi kinerja model yang objektif [15], [16].

### C. Ketidakseimbangan Kelas

Ketidakseimbangan kelas (class imbalance) merupakan kondisi ketika distribusi jumlah sampel antar kelas target tidak merata, sehingga satu kelas memiliki dominasi yang signifikan dibandingkan kelas lainnya. Fenomena ini umum dijumpai pada dataset material, termasuk perovskite oksida, dan merefleksikan distribusi alami ruang komposisi material yang dipengaruhi oleh stabilitas termodinamika serta preferensi ikatan kimia tertentu [8]. Akibatnya, konfigurasi struktur dan sifat elektronik tertentu, termasuk jenis band gap, muncul lebih sering dibandingkan konfigurasi lainnya [17].

Dalam konteks klasifikasi jenis band gap, kondisi ini menyebabkan representasi data yang tidak proporsional antara kelas direct dan indirect band gap [17].

Ketidakseimbangan kelas berdampak langsung terhadap kinerja model klasifikasi, di mana model cenderung memprioritaskan kelas mayoritas dalam proses pembelajaran, sehingga menghasilkan decision boundary yang kurang representatif dan menurunkan kemampuan model dalam mengenali kelas minoritas, meskipun nilai akurasi keseluruhan tampak tinggi [18]. Permasalahan ini semakin kompleks ketika terjadi tumpang tindih distribusi fitur antar kelas, yang berpotensi menimbulkan bias prediksi dan menurunkan kemampuan generalisasi model terhadap kandidat material baru [18], [19], [20]. Oleh karena itu, penanganan ketidakseimbangan kelas menjadi aspek metodologis yang krusial dalam pengembangan model klasifikasi sifat material berbasis Machine Learning [20].

### D. Penanganan Ketidakseimbangan Kelas dengan Tomek Links

Ketidakseimbangan kelas merupakan permasalahan umum dalam klasifikasi sifat material berbasis Machine Learning [21] yang berpotensi menyebabkan bias model terhadap kelas mayoritas serta menurunkan kemampuan deteksi kelas minoritas [19], [20]. Oleh karena itu, diperlukan strategi penanganan ketidakseimbangan kelas yang selaras dengan karakteristik data dan tujuan klasifikasi.

Dalam penelitian ini, metode Tomek Links diterapkan sebagai pendekatan undersampling berbasis batas keputusan (decision boundary) untuk mengurangi tumpang tindih distribusi fitur antar kelas [12], [22], [23]. Tomek Links mengidentifikasi pasangan sampel dari kelas berbeda yang saling menjadi nearest neighbor, yang merepresentasikan wilayah overlap dan potensi ambiguitas dalam proses klasifikasi [12], [22].

Penerapan Tomek Links dilakukan dengan menghapus sampel dari kelas mayoritas yang terlibat dalam pasangan Tomek Links, sementara sampel dari kelas minoritas tetap dipertahankan [12], [22], [23]. Metode ini diterapkan secara eksklusif pada data latih setelah proses pembagian data dan sebelum tahap standarisasi fitur, guna mencegah terjadinya data leakage dan menjaga validitas evaluasi model [18]. Setelah proses undersampling, fitur numerik pada data latih distandarasi, dan parameter transformasi yang sama diterapkan pada data uji untuk menjaga konsistensi skala antar partisi data. Selanjutnya, model dievaluasi menggunakan stratified k-fold cross-validation pada data latih, sementara evaluasi performa akhir dilakukan pada data uji terpisah dengan metrik yang sesuai untuk kondisi data tidak seimbang.

### E. Model Klasifikasi

Penelitian ini menggunakan beberapa algoritma klasifikasi untuk menganalisis pengaruh penerapan Tomek Links pada kondisi data tidak seimbang, yaitu Multi-Layer Perceptron

(MLP), CatBoost, Extra Trees Classifier, dan Gradient Boosting. Pemilihan model didasarkan pada perbedaan karakteristik pembelajaran, khususnya dalam menangkap hubungan nonlinier, ketahanan terhadap noise, serta sensitivitas terhadap tumpang tindih distribusi fitur pada data numerik.

Model MLP dibangun dengan dua hidden layer berukuran 64 dan 32 neuron, menggunakan fungsi aktivasi ReLU dan optimizer Adam dengan maksimum 500 iterasi. Model ensemble lainnya digunakan dengan parameter bawaan (default) dari pustaka yang digunakan, dengan pengaturan random state yang konsisten untuk menjaga keterulangan dan keadilan perbandingan performa antar model.

Sebelum pemodelan, seluruh fitur numerik distandarisasi menggunakan StandardScaler. Pelatihan dan validasi model dilakukan pada data latih menggunakan skema Stratified K-Fold Cross-Validation dengan lima lipatan, sedangkan evaluasi performa akhir dilakukan pada data uji terpisah untuk menilai kemampuan generalisasi model.

#### F. Skema Validasi dan Evaluasi

Kinerja klasifikasi dievaluasi menggunakan beberapa metrik, yaitu accuracy, Precision, Recall, F1-score dan confusion matrix. Meskipun akurasi tetap dilaporkan sebagai indikator performa global, fokus utama evaluasi diberikan pada F1-score makro karena dataset memiliki distribusi kelas yang tidak seimbang [24], [25], [26].

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

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

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

$$F1 \text{ Score} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Keterangan:

TP : True Positive

TN : True Negative

FP : False Positive

FN : False Negative

Penggunaan F1-score makro memungkinkan penilaian performa yang lebih adil dengan memberikan bobot yang setara pada setiap kelas, sehingga lebih representatif dibandingkan akurasi dalam kondisi ketidakseimbangan kelas [27], [28]. Hasil evaluasi pada tahap pelatihan dilaporkan sebagai rata-rata dari Stratified K-Fold Cross-Validation, sehingga setiap fold mempertahankan proporsi kelas asli dan memungkinkan penilaian performa model yang lebih stabil [29], [30], [31]. Evaluasi akhir dilakukan menggunakan data uji terpisah, untuk menilai kemampuan generalisasi model

terhadap data baru dan memastikan performa tidak hanya mencerminkan kecocokan terhadap data latih [32], [33], [34].

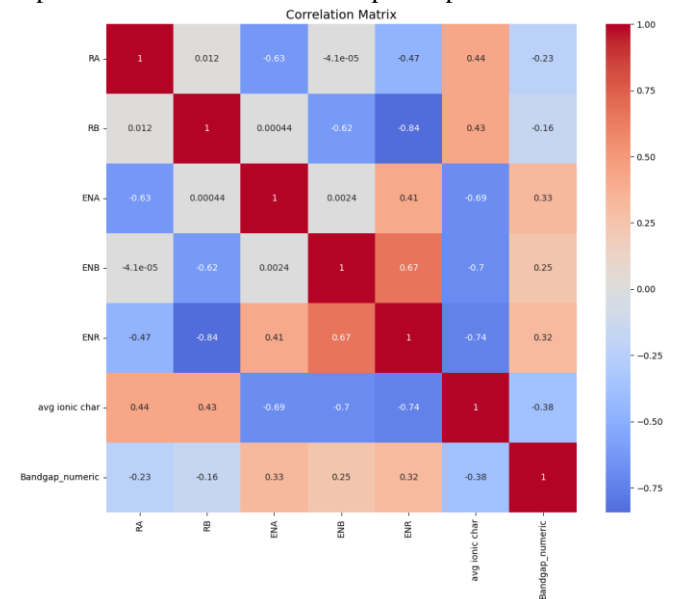
### III. HASIL DAN PEMBAHASAN

#### A. Deskripsi Dataset dan Kondisi Ketidakseimbangan Kelas

Dataset perovskite oksida  $ABO_3$  yang digunakan dalam penelitian ini pada tahap awal terdiri dari 5.940 sampel dengan 7 atribut, yang mencakup enam fitur numerik sebagai variabel prediktor dan satu variabel target berupa kelas jenis band gap (direct dan indirect). Pada kondisi awal, distribusi kelas target bersifat seimbang dengan rasio 50:50.

Setelah dilakukan proses pembersihan data melalui penghapusan duplikat, jumlah sampel berkurang menjadi 3.469 data. Distribusi kelas pada dataset hasil pembersihan menunjukkan ketidakseimbangan yang signifikan, dengan 2.913 sampel (84%) pada kelas direct dan 556 sampel (16%) pada kelas indirect. Ketidakseimbangan ini berpotensi menyebabkan bias model terhadap kelas mayoritas serta menurunkan sensitivitas terhadap kelas minoritas.

Analisis matriks korelasi (Gambar 2) menunjukkan adanya korelasi yang relatif tinggi pada beberapa pasangan fitur, yang mengindikasikan potensi redundansi informasi. Selain itu, distribusi data antar kelas pada ruang fitur memperlihatkan adanya kemungkinan overlap pada area batas keputusan. Kondisi ini mendasari penerapan metode Tomek Links untuk membersihkan sampel borderline dan meningkatkan separabilitas antar kelas sebelum proses pemodelan.



Gambar 2. Correlation Matrix

#### B. Dampak Penerapan Tomek Links terhadap Distribusi Data

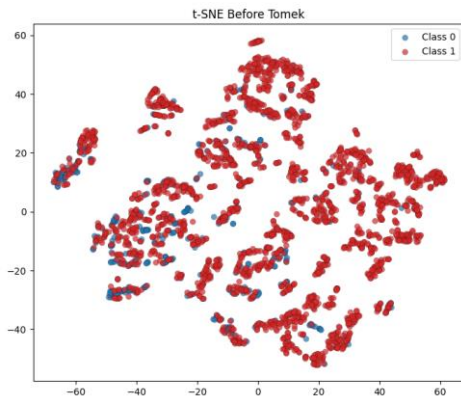
Penerapan metode Tomek Links dilakukan secara eksklusif pada data latih untuk mengurangi tumpang tindih distribusi fitur antar kelas di sekitar batas keputusan tanpa memengaruhi representasi kelas minoritas. Sebelum penerapan, data latih terdiri atas 2.775 sampel dengan 6 fitur

numerik, dengan distribusi 2.330 sampel pada kelas direct dan 445 sampel pada kelas indirect, yang setara dengan proporsi 84:16.

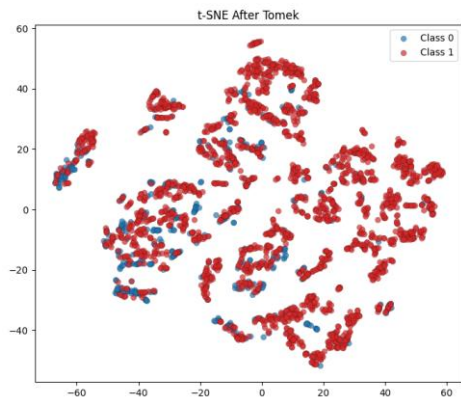
Setelah penerapan Tomek Links, jumlah sampel berkurang menjadi 2.639. Seluruh 136 sampel yang dihapus (4,90%) berasal dari kelas mayoritas, sehingga jumlah kelas mayoritas berkurang menjadi 2.194, sementara kelas minoritas tetap 445 sampel. Proporsi ketidakseimbangan kelas berubah relatif kecil dari 84:16 menjadi 83:17. Temuan ini menunjukkan bahwa Tomek Links tidak berfungsi sebagai teknik rebalancing, melainkan sebagai mekanisme reduksi ambiguitas lokal pada area batas keputusan.

TABEL II  
PERBANDINGAN DISTRIBUSI DATA TRAIN SEBELUM DAN SESUDAH TOMEK LINK

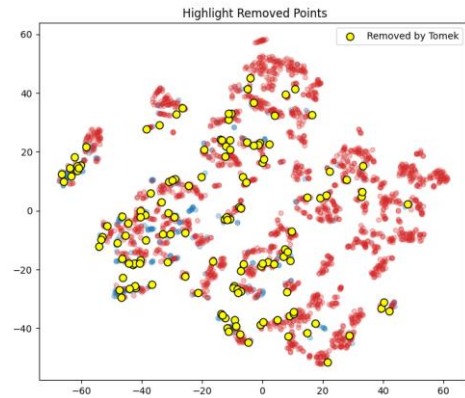
Kondisi Data	Sebelum Tomek Links	Sesudah Tomek Links
Jumlah Sample	2.775	2.639
Direct (1)	2.330	2.194
Indirect (0)	445	445
Rasio Kelas	84:16	83:17



Gambar 3. Visualisasi data sebelum tomeklinks menggunakan t-SNE



Gambar 4. Visualisasi data sesudah tomeklinks menggunakan t-SNE



Gambar 5. Visualisasi data sebelum tomeklinks disertai highlight removed points menggunakan t-SNE

Untuk mengevaluasi distribusi spasial sampel yang dieliminasi, dilakukan visualisasi reduksi dimensi menggunakan t-SNE (Gambar 3–5). Hasil visual menunjukkan bahwa sampel yang dihapus terkonsentrasi pada zona transisi antar kelas (*overlap region*), sementara inti klaster masing-masing kelas tetap terjaga. Meskipun t-SNE tidak sepenuhnya mempertahankan struktur global, pola distribusi lokal yang diamati mendukung bahwa proses pembersihan terjadi pada area dengan kedekatan lintas kelas yang tinggi.

Dengan proporsi penghapusan yang relatif rendah dan tanpa reduksi pada kelas minoritas, risiko hilangnya informasi diskriminatif dinilai terbatas. Sebaliknya, pembersihan ini berpotensi meningkatkan kejelasan margin antar kelas dengan meminimalkan pengaruh noise pada area boundary.

### C. Performa Model Tanpa Penanganan Ketidakseimbangan Kelas

Evaluasi baseline dilakukan tanpa menerapkan teknik penanganan ketidakseimbangan kelas untuk mengukur kemampuan intrinsik masing-masing model dalam menghadapi distribusi kelas yang timpang. Berdasarkan hasil pengujian pada data uji, seluruh model menghasilkan nilai akurasi yang relatif tinggi, yaitu berada pada rentang 0,8703 hingga 0,8934 sebagaimana ditunjukkan pada Tabel III. Model MLP memperoleh akurasi tertinggi sebesar 0,8934, diikuti oleh CatBoost (0,8847), Gradient Boosting (0,8833), dan Extra Trees (0,8703).

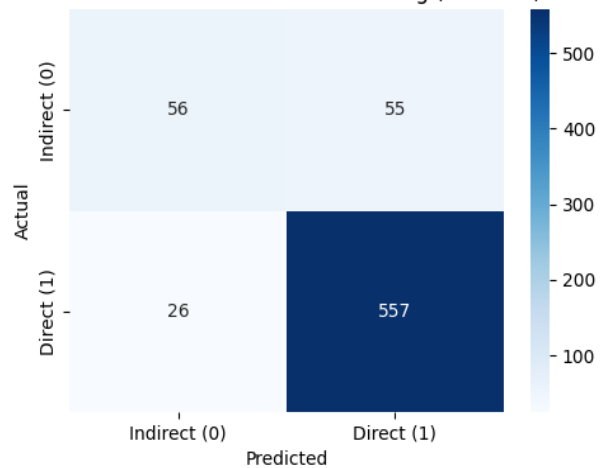
TABEL III  
PERFORMA MODEL BASELINE

Model	Accuracy	Macro F1	Recall (Indirect 0)	Recall (Direct)
MLP	0,8934	0,7783	0,54	0,96
CatBoost	0,8847	0,7603	0,51	0,96
Gradient Boosting	0,8833	0,7563	0,50	0,96
Extra Trees	0,8703	0,7349	0,49	0,94

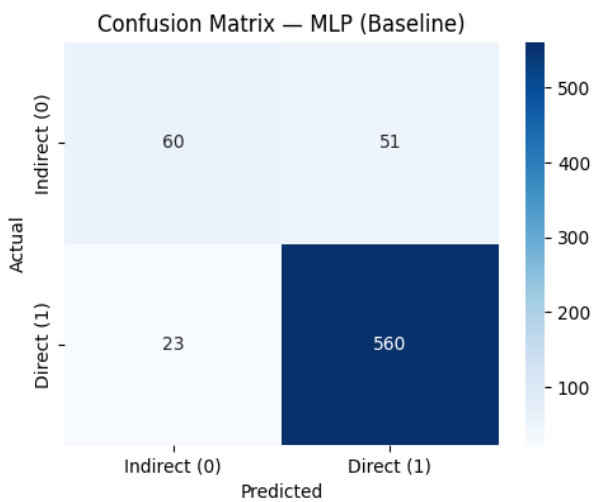
Namun demikian, evaluasi menggunakan macro F1-score menunjukkan nilai yang lebih rendah, yaitu berada pada rentang 0,7349 hingga 0,7783. Nilai macro F1 tertinggi diperoleh oleh MLP sebesar 0,7783, sedangkan nilai terendah terdapat pada Extra Trees sebesar 0,7349. Perbedaan antara akurasi dan macro F1-score ini mengindikasikan bahwa performa model tidak seimbang antar kelas, terutama pada kelas minoritas (Indirect).

Analisis lebih lanjut terhadap nilai recall menunjukkan adanya kesenjangan sensitivitas yang signifikan antara kedua kelas. Recall untuk kelas Indirect (0) berada pada kisaran 0,49 hingga 0,54, sedangkan recall kelas Direct (1) konsisten tinggi pada kisaran 0,94 hingga 0,96. Pada model MLP, dari total 111 sampel aktual kelas Indirect, hanya sekitar 54% atau 59 hingga 60 sampel yang berhasil diklasifikasikan dengan benar, sementara sisanya salah diprediksi sebagai kelas Direct. Kondisi ini lebih rendah pada model Extra Trees dengan recall sebesar 0,49, yang berarti lebih dari setengah sampel kelas minoritas gagal terdeteksi.

Confusion Matrix — GradientBoosting (Baseline)

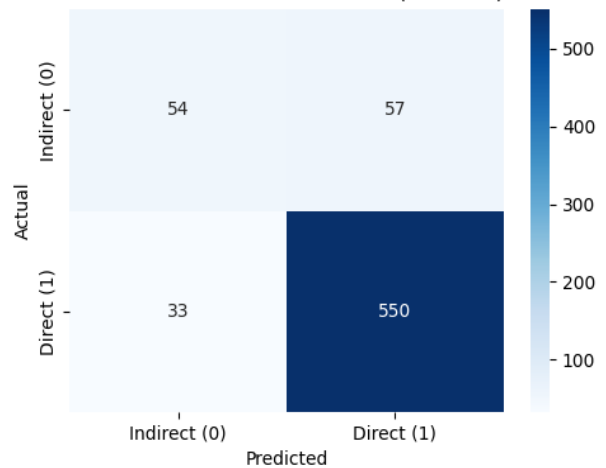


Gambar 7. Confusion matrix Gradient Boosting sebelum Tomek Link

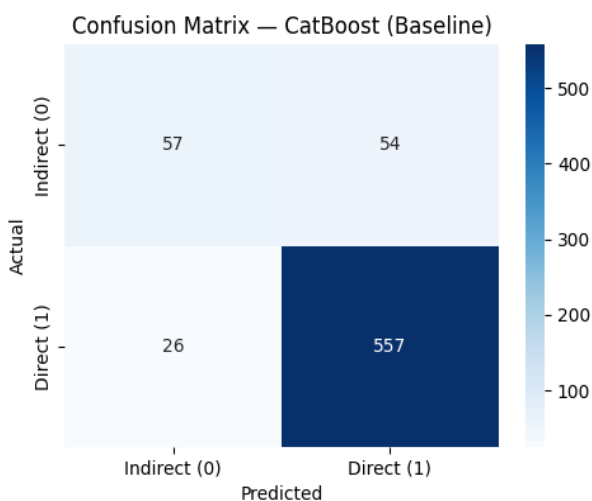


Gambar 5. Confusion matrix MLP sebelum Tomek Link

Confusion Matrix — ExtraTrees (Baseline)



Gambar 8. Confusion matrix Extra Trees sebelum Tomek Link



Gambar 6. Confusion matrix CatBoost sebelum Tomek Link

Hasil 5-Fold Stratified Cross-Validation pada eksperimen baseline disajikan pada Tabel IV. Evaluasi ini bertujuan untuk mengukur konsistensi performa model pada berbagai partisi data latih tanpa penerapan teknik penanganan ketidakseimbangan kelas.

Berdasarkan hasil validasi silang, model Gradient Boosting memperoleh rata-rata akurasi tertinggi sebesar 0,8890 dengan interval kepercayaan 95% pada rentang [0,8845–0,8935], diikuti oleh MLP sebesar 0,8879 [0,8852–0,8905], CatBoost sebesar 0,8864 [0,8812–0,8917], dan Extra Trees sebesar 0,8732 [0,8666–0,8797].

Dari sisi macro F1-score, MLP menunjukkan performa terbaik dengan nilai rata-rata 0,7593 [0,7411–0,7775], diikuti oleh Gradient Boosting sebesar 0,7553 [0,7461–0,7645], CatBoost sebesar 0,7507 [0,7377–0,7637], dan Extra Trees sebesar 0,7367 [0,7233–0,7502].

TABEL IV  
HASIL 5-FOLD STRATIFIED CROSS-VALIDATION (BASELINE)

Model	Accuracy (Mean ± 95% CI)	Macro F1 (Mean ± 95% CI)
MLP	0,8879 ± 0,0019	0,7593 ± 0,0131
CatBoost	0,8864 ± 0,0038	0,7507 ± 0,0094
Gradient Boosting	0,8890 ± 0,0032	0,7553 ± 0,0066
Extra Trees	0,8732 ± 0,0047	0,7367 ± 0,0097

Interval kepercayaan yang relatif sempit pada seluruh model mengindikasikan bahwa performa yang diamati stabil di berbagai fold dan bukan merupakan hasil fluktuasi acak akibat pembagian data tertentu. Meskipun demikian, nilai macro F1 yang konsisten lebih rendah dibandingkan akurasi menegaskan bahwa keterbatasan dalam mendeteksi kelas minoritas bersifat sistematis pada kondisi tanpa penanganan ketidakseimbangan kelas.

Berdasarkan temuan tersebut, diperlukan pendekatan yang secara eksplisit menangani distribusi kelas yang tidak seimbang untuk meningkatkan sensitivitas model terhadap kelas minoritas tanpa mengorbankan stabilitas prediksi kelas mayoritas.

*D. Performa Model dengan Penerapan Tomek Links*

Evaluasi performa model setelah penerapan Tomek Links dilakukan menggunakan prosedur yang identik dengan eksperimen baseline, yaitu pelatihan pada data latih, pengujian pada data uji terpisah, serta validasi menggunakan 5-Fold Stratified Cross-Validation. Pendekatan ini memastikan bahwa setiap perubahan performa yang diamati dapat dikaitkan secara langsung dengan mekanisme penanganan ketidakseimbangan kelas, bukan akibat perbedaan skema evaluasi.

TABEL V  
PERFORMA MODEL DENGAN TOMEK LINKS

Model	Accuracy	Macro F1	Recall (Indirect 0)	Recall (Direct)
MLP	0,8905	0,7761	0,55	0,96
CatBoost	0,8804	0,7681	0,58	0,94
Gradient Boosting	0,8847	0,7756	0,59	0,94
Extra Trees	0,8631	0,7462	0,58	0,92

Berdasarkan hasil pengujian pada data uji, performa model setelah penerapan Tomek Links disajikan pada Tabel V. Nilai akurasi berada pada rentang 0,8631 hingga 0,8905. Model MLP memperoleh akurasi tertinggi sebesar 0,8905, diikuti oleh Gradient Boosting (0,8847), CatBoost (0,8804), dan Extra Trees (0,8631). Rentang ini relatif sebanding dengan baseline, yang menunjukkan bahwa penerapan Tomek Links tidak ditujukan untuk meningkatkan akurasi global secara

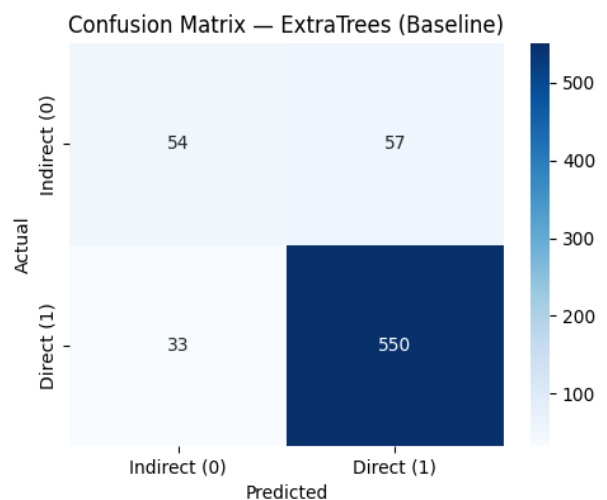
substansial, melainkan untuk memperbaiki keseimbangan performa antar kelas.

Perubahan yang lebih relevan terlihat pada metrik macro F1-score, yang berada pada kisaran 0,7462 hingga 0,7761. Nilai tertinggi diperoleh oleh MLP sebesar 0,7761, diikuti sangat dekat oleh Gradient Boosting sebesar 0,7756. Jika dibandingkan dengan baseline, terjadi peningkatan macro F1 sekitar 0,01–0,02 poin pada sebagian besar model. Meskipun peningkatan ini bersifat moderat, pola tersebut mengindikasikan perbaikan distribusi performa antar kelas, terutama dalam konteks data tidak seimbang.

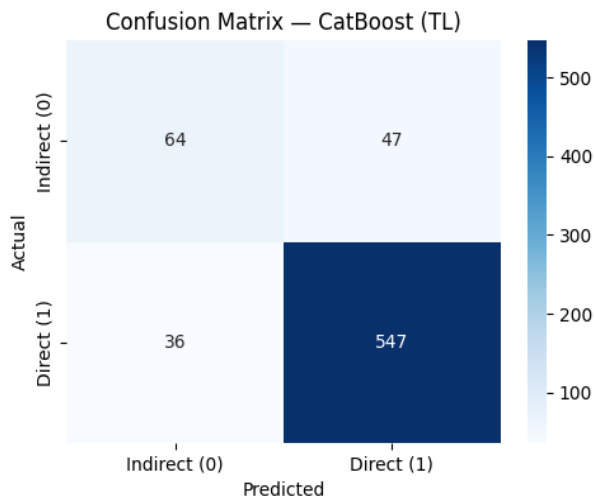
Analisis recall menunjukkan adanya peningkatan sensitivitas terhadap kelas minoritas (Indirect). Nilai recall meningkat ke kisaran 0,55–0,59 dibandingkan baseline yang berada pada rentang 0,49–0,54. Dengan total 111 sampel kelas Indirect pada data uji, recall sebesar 0,59 pada Gradient Boosting berarti sekitar 65 sampel berhasil diklasifikasikan dengan benar, meningkat dari sekitar 55–60 sampel pada kondisi tanpa penanganan ketidakseimbangan. Peningkatan ini menunjukkan bahwa Tomek Links mampu mengurangi kesalahan klasifikasi berupa false negative pada kelas minoritas.

Sebaliknya, recall kelas mayoritas (Direct) mengalami sedikit penurunan pada beberapa model, dari kisaran 0,94–0,96 menjadi 0,92–0,94. Penurunan ini relatif kecil dan mencerminkan adanya trade-off antara akurasi global dan keseimbangan performa antar kelas. Dengan kata lain, sebagian prediksi yang sebelumnya terdistribusi secara dominan pada kelas mayoritas kini lebih seimbang sehingga meningkatkan deteksi kelas minoritas.

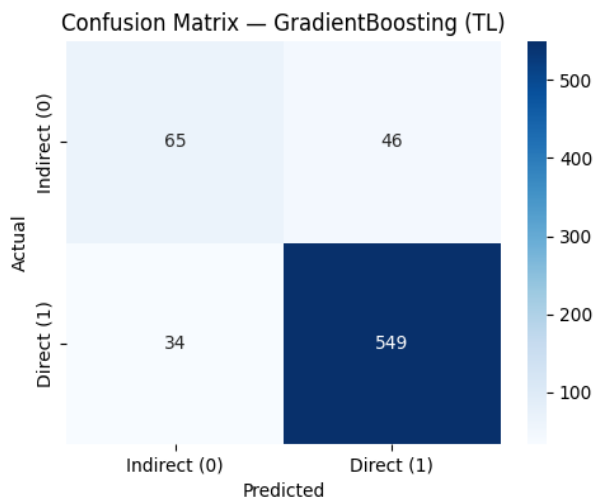
Temuan tersebut diperkuat oleh visualisasi confusion matrix pada Gambar 9 hingga Gambar 12, yang menunjukkan penurunan jumlah false negative pada kelas Indirect dibandingkan kondisi baseline. Meskipun terjadi sedikit peningkatan false positive terhadap kelas mayoritas, distribusi kesalahan menjadi lebih proporsional antar kelas.



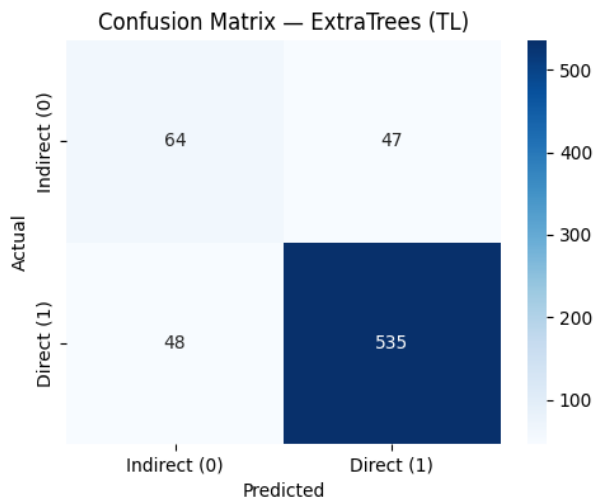
Gambar 9. Confusion matrix MLP setelah Tomek Link



Gambar 10. Confusion matrix CatBoost setelah Tomek Link



Gambar 11. Confusion matrix Gradient Boosting setelah Tomek Link



Gambar 12. Confusion matrix Extra Trees setelah Tomek Link

Dari sisi generalisasi, model Extra Trees menunjukkan kesenjangan yang besar antara performa pelatihan dan

pengujian (accuracy train 0,9989 dan test 0,8631), yang mengindikasikan kecenderungan overfitting meskipun recall minoritas meningkat. Sebaliknya, MLP dan Gradient Boosting memperlihatkan selisih train dan test yang lebih terkendali, menunjukkan stabilitas model terhadap perubahan distribusi akibat proses decision boundary cleaning.

Hasil 5-Fold Stratified Cross-Validation setelah penerapan Tomek Links disajikan pada Tabel VI. Evaluasi ini digunakan untuk mengukur konsistensi performa pada berbagai partisi data latih.

TABEL VI  
HASIL 5-FOLD STRATIFIED CROSS-VALIDATION

Model	Accuracy (Mean ± 95% CI)	Macro F1 (Mean ± 95% CI)
MLP	0,8844 [0,8778–0,8910]	0,7682 [0,7483–0,7880]
CatBoost	0,8818 [0,8730–0,8906]	0,7556 [0,7313–0,7800]
Gradient Boosting	0,8870 [0,8825–0,8915]	0,7627 [0,7478–0,7776]
Extra Trees	0,8660 [0,8528–0,8791]	0,7373 [0,7132–0,7614]

Interval kepercayaan yang relatif sempit pada seluruh model menunjukkan bahwa peningkatan performa terhadap kelas minoritas bersifat konsisten pada berbagai fold dan bukan merupakan fluktuasi acak akibat pembagian data tertentu. MLP dan Gradient Boosting menunjukkan stabilitas yang lebih baik dibandingkan Extra Trees, yang memiliki variabilitas lebih besar sejalan dengan kecenderungan overfitting pada data latih.

Secara keseluruhan, penerapan Tomek Links tidak menghasilkan peningkatan akurasi global yang besar, namun secara konsisten meningkatkan recall kelas minoritas dan macro F1-score. Temuan ini menegaskan bahwa pendekatan decision boundary cleaning lebih efektif dalam mengurangi bias terhadap kelas mayoritas dibandingkan dalam meningkatkan akurasi total, sehingga relevan untuk permasalahan klasifikasi dengan distribusi kelas yang tidak seimbang.

*E. Analisis Perbandingan Sebelum dan Sesudah Tomek Links*

Analisis komparatif antara eksperimen baseline dan model dengan penerapan Tomek Links menunjukkan bahwa perubahan performa yang dihasilkan bersifat selektif dan terutama terfokus pada peningkatan kinerja kelas minoritas. Secara umum, penerapan Tomek Links menghasilkan perubahan akurasi yang relatif kecil (−0,72% hingga +0,14%), sementara recall kelas Indirect meningkat secara konsisten pada seluruh model. Peningkatan terbesar mencapai +9% pada Gradient Boosting dan Extra Trees. Kenaikan sensitivitas terhadap kelas minoritas ini berdampak langsung pada peningkatan macro F1-score, yang mengonfirmasi bahwa Tomek Links mampu memperbaiki keseimbangan performa antar kelas dengan trade-off yang terbatas pada kelas mayoritas.

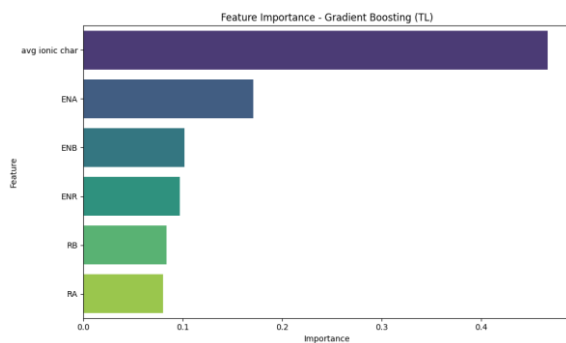
Ditinjau dari karakteristik masing-masing model, Gradient Boosting dan CatBoost menunjukkan trade-off yang paling seimbang, dengan peningkatan recall kelas Indirect yang signifikan tanpa penurunan performa global yang substansial. MLP memperlihatkan perubahan yang relatif kecil, yang mengindikasikan bahwa model ini telah memiliki decision boundary yang cukup stabil sejak kondisi baseline. Sebaliknya, Extra Trees memperoleh peningkatan terbesar pada recall kelas minoritas, namun disertai penurunan akurasi dan recall kelas mayoritas yang lebih nyata. Pola ini mengindikasikan bahwa Extra Trees lebih sensitif terhadap perubahan distribusi lokal data latih.

Secara konseptual, temuan ini menegaskan bahwa efektivitas Tomek Links bersifat model-dependent dan paling optimal ketika diterapkan pada model yang sensitif terhadap overlap fitur di sekitar decision boundary. Dengan demikian, Tomek Links berfungsi lebih sebagai mekanisme decision boundary cleaning dibandingkan sebagai metode penyeimbangan kelas berbasis manipulasi kuantitas data semata.

TABEL V II  
PERUBAHAN PERFORMA MODEL ( $\Delta$ ) SETELAH PENERAPAN  
TOMEK LINK

Model	$\Delta$ Accuracy	$\Delta$ Macro F1	$\Delta$ Recall Indirect $\Delta$ (0)	$\Delta$ Recall Direct (1)
MLP	-0,0029	-0,0022	+0,01	0,00
CatBoost	-0,0043	+0,0078	+0,07	-0,02
Gradient Boosting	+0,0014	+0,0193	+0,09	-0,02
Extra Trees	-0,0072	+0,0113	+0,09	-0,02

#### F. Analisis Kontribusi Fitur terhadap Klasifikasi Band Gap



Gambar 12. Feature Important dari model terbaik setelah Tomek Link

Untuk mengevaluasi apakah model tidak hanya memberikan performa prediktif yang tinggi tetapi juga memiliki keterkaitan dengan prinsip fisika material, dilakukan analisis kontribusi fitur menggunakan model terbaik, yaitu Gradient Boosting setelah penerapan Tomek Links. Analisis ini bertujuan untuk mengidentifikasi fitur yang paling berpengaruh dalam membedakan material

dengan karakter band gap direct dan indirect, serta menilai keselarasan hasil model dengan teori struktur pita.

Berdasarkan nilai feature importance yang dihasilkan, fitur dengan kontribusi terbesar adalah average ionic character (avg ionic char), dengan nilai kepentingan yang secara signifikan lebih tinggi dibandingkan fitur lainnya ( $\approx 0,47$ ). Dominasi fitur ini menunjukkan bahwa tingkat karakter ionik dalam ikatan kimia merupakan determinan utama dalam proses klasifikasi. Secara fisika, karakter ionik memengaruhi distribusi muatan dan sifat orbital yang terlibat dalam pembentukan pita valensi dan pita konduksi. Material dengan perbedaan elektronegativitas yang besar cenderung memiliki kontribusi ionik yang lebih tinggi, yang dapat memengaruhi bentuk kurva dispersi energi dan lokasi minimum pita konduksi relatif terhadap maksimum pita valensi [35], [36]. Hal ini relevan terhadap kemungkinan terjadinya transisi direct maupun indirect pada struktur pita.

Sementara itu, fitur RB dan RA yang berkaitan dengan parameter jari-jari atom memiliki kontribusi relatif lebih rendah, namun tetap berperan dalam membentuk geometri kristal dan parameter kisi. Perbedaan ukuran atom dapat memengaruhi jarak antar atom dalam kisi kristal, yang berdampak pada tumpang tindih orbital dan bentuk dispersi pita energi [37]. Meskipun kontribusinya tidak sebesar karakter ionik dan elektronegativitas, fitur geometris ini tetap mendukung kemampuan model dalam memisahkan kedua kelas.

Menariknya, dominasi avg ionic char yang konsisten dengan teori kimia ikatan menunjukkan bahwa model tidak semata-mata menangkap pola statistik dalam data, melainkan juga merefleksikan hubungan fisika yang mendasari pembentukan band gap. Selain itu, penerapan Tomek Links tidak menggeser struktur kontribusi fitur secara drastis, yang mengindikasikan bahwa metode tersebut berfungsi terutama dalam membersihkan batas keputusan antar kelas tanpa mengubah hubungan fundamental antara fitur dan label. Dengan demikian, peningkatan performa model setelah penerapan Tomek Links dapat diinterpretasikan sebagai hasil dari perbaikan kualitas decision boundary, bukan akibat perubahan distribusi informasi intrinsik pada fitur dominan.

Secara keseluruhan, analisis ini memperkuat validitas ilmiah model yang dikembangkan, karena fitur-fitur yang paling berpengaruh memiliki dasar fisika yang jelas dalam teori struktur pita dan mekanisme transisi elektronik pada material semikonduktor.

#### G. Implikasi Hasil Klasifikasi pada Studi Material

Peningkatan performa klasifikasi pada kelas minoritas setelah penerapan Tomek Links memiliki implikasi penting dalam studi material berbasis data. Dalam klasifikasi sifat elektronik seperti band gap, peningkatan recall kelas minoritas menurunkan risiko kesalahan false negative yang dapat menyebabkan kandidat material potensial tereliminasi pada tahap awal material discovery maupun high-throughput screening.

Dengan berkurangnya bias terhadap kelas mayoritas, model menjadi lebih sensitif dalam mengidentifikasi material dengan karakteristik elektronik tertentu yang jumlahnya relatif lebih sedikit dalam dataset. Hal ini meningkatkan reliabilitas model sebagai alat bantu pengambilan keputusan dalam studi material berbasis pembelajaran mesin, terutama pada skenario eksplorasi material baru di mana kegagalan mendeteksi kandidat potensial dapat berdampak signifikan terhadap efisiensi proses penelitian.

#### H. Batasan Penelitian dan Generalisasi Model

Penelitian ini berfokus secara spesifik pada keluarga perovskite oksida dengan struktur kristal  $ABO_3$ . Meskipun model menunjukkan performa yang andal pada dataset ini, perlu dicatat bahwa generalisasi model terhadap keluarga perovskite lain, seperti perovskite halida atau double perovskite, belum diuji secara eksternal dalam studi ini. Perbedaan dalam kation A dan B serta konfigurasi oktahedral oksigen pada  $ABO_3$  memberikan karakteristik elektronik yang unik, sehingga penerapan model pada struktur kristal yang berbeda mungkin memerlukan fine-tuning atau penambahan fitur deskriptor baru yang relevan dengan kimiawi material tersebut.

#### IV. KESIMPULAN

Penelitian ini mengevaluasi pengaruh penerapan Tomek Links terhadap performa klasifikasi jenis band gap (direct dan indirect) pada material perovskite oksida  $ABO_3$  dengan ketidakseimbangan kelas yang signifikan. Analisis awal menunjukkan bahwa model baseline cenderung bias terhadap kelas mayoritas, dengan akurasi global tinggi namun recall dan macro F1-score rendah pada kelas minoritas (indirect). Temuan ini menegaskan bahwa akurasi saja tidak cukup untuk menilai performa model pada dataset yang tidak seimbang.

Penerapan Tomek Links pada data latih terbukti efektif dalam membersihkan decision boundary melalui penghapusan selektif sampel mayoritas di area overlap antar kelas, tanpa mengurangi representasi kelas minoritas. Proses ini meningkatkan kejelasan margin antar kelas dan mengurangi ambiguitas lokal, sehingga mendukung pembelajaran model yang lebih seimbang. Visualisasi t-SNE menunjukkan bahwa penghapusan terjadi di zona transisi, sementara inti klaster masing-masing kelas tetap terjaga.

Hasil evaluasi menunjukkan bahwa akurasi global tetap stabil (86,31%–89,05%), namun secara konsisten terjadi peningkatan recall kelas minoritas dan macro F1-score pada seluruh model. Model berbasis boosting, khususnya Gradient Boosting dan CatBoost, menunjukkan trade-off performa paling optimal, dengan peningkatan sensitivitas minoritas tanpa penurunan performa global yang berarti. Model lain seperti Extra Trees menunjukkan peningkatan recall minoritas, namun disertai penurunan akurasi dan recall kelas mayoritas yang lebih nyata, menegaskan bahwa efektivitas Tomek Links bersifat model-dependent.

Secara keseluruhan, metode Tomek Links berperan sebagai mekanisme decision boundary cleaning yang efektif untuk klasifikasi sifat elektronik material berbasis Machine Learning pada dataset tidak seimbang. Peningkatan sensitivitas terhadap kelas minoritas memiliki implikasi penting dalam material informatics dan high-throughput screening, karena mengurangi risiko false negative dan meningkatkan reliabilitas model sebagai alat bantu eksplorasi material berbasis data. Penelitian lanjutan dapat mengeksplorasi penerapan model pada keluarga perovskite lain atau kombinasi dengan teknik penyeimbangan tambahan untuk memperluas generalisasi dan stabilitas performa.

#### DAFTAR PUSTAKA

- [1] S. Rahman *et al.*, "Structural, electronic, optical and mechanical properties of oxide-based perovskite  $ABO_3$  (A = Cu, Nd and B = Sn, Se): A DFT study," *J. Solid State Chem.*, vol. 317, p. 123650, Jan. 2023, doi: 10.1016/j.jssc.2022.123650.
- [2] Y. Choi *et al.*, "Perovskite nanocomposites: synthesis, properties, and applications from renewable energy to optoelectronics," Dec. 01, 2024, *Korea Nano Technology Research Society*. doi: 10.1186/s40580-024-00440-7.
- [3] S. Rahman *et al.*, "Exploring the multifaceted properties of novel oxide-based perovskites  $ABO_3$  (A=Nd and BLr, Y): A DFT study," *Mater. Sci. Semicond. Process.*, vol. 180, p. 108558, Sep. 2024, doi: 10.1016/j.mssp.2024.108558.
- [4] M. Fatmi *et al.*, "Structural, electronic, optical, and thermoelectric properties of  $CaXO_3$  (X = Si, Ge, Ti) perovskite for photovoltaics and optical devices," *Sci. Rep.*, vol. 15, no. 1, Dec. 2025, doi: 10.1038/s41598-025-31002-4.
- [5] F. Dinic, I. Neporozhnyi, and O. Voznyy, "Machine learning models for the discovery of direct band gap materials for light emission and photovoltaics," *Comput. Mater. Sci.*, vol. 231, p. 112580, Jan. 2024, doi: 10.1016/j.commatsci.2023.112580.
- [6] A. Sabagh Moeini, F. Shariatmadar Tehrani, and A. Naeimi-Sadigh, "Machine learning-enhanced band gaps prediction for low-symmetry double and layered perovskites," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-77081-7.
- [7] F. J. Kusuma *et al.*, "Direct band gap prediction of single and double perovskite using cost-sensitive ensemble learning," *J. Alloys Compd.*, vol. 1037, p. 182102, Aug. 2025, doi: 10.1016/j.jallcom.2025.182102.
- [8] S. P. G. M. N. Matur, N. Nagappan, S. Rath, and T. Thomas, "Prediction of nature of band gap of perovskite oxides ( $ABO_3$ ) using a machine learning approach," *Journal of Materiomics*, vol. 8, no. 5, pp. 937–948, Sep. 2022, doi: 10.1016/j.jmat.2022.04.006.
- [9] J. Jiang *et al.*, "A review of machine learning methods for imbalanced data challenges in chemistry," Apr. 22, 2025, *Royal Society of Chemistry*. doi: 10.1039/d5sc00270b.
- [10] A. Mathew, A. A. B. Baloch, A. Yakasai, H. Mittal, V. Alberts, and J. V. Karunamurthy, "Machine learning-driven crystal system prediction for perovskites using augmented XRD data," *Eng. Appl. Artif. Intell.*, vol. 164, p. 113247, Jan. 2026, doi: 10.1016/j.engappai.2025.113247.
- [11] D. Devi, S. kr Biswas, and B. Purkayastha, "Redundancy-driven modified Tomek-link based undersampling: A solution to class imbalance," *Pattern Recognit. Lett.*, vol. 93, pp. 3–12, Jul. 2017, doi: 10.1016/j.patrec.2016.10.006.
- [12] E. F. Swana, W. Doorsamy, and P. Bokoro, "Tomek Link and SMOTE Approaches for Machine Fault Classification with an Imbalanced Dataset," *Sensors*, vol. 22, no. 9, May 2022, doi: 10.3390/s22093246.
- [13] E. Ogoshi *et al.*, "Learning from machine learning: the case of band-gap directness in semiconductors," *Discover Materials*, vol. 4, no. 1, Dec. 2024, doi: 10.1007/s43939-024-00073-x.

- [14] Y. Oh, S. Song, and J. Bae, "A Review of Bandgap Engineering and Prediction in 2D Material Heterostructures: A DFT Perspective," Dec. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/ijms252313104.
- [15] N. Prameswari, W. Ghozi, and F. A. Rafrastara, "Systematic XGBoost Pipeline for Phishing Website Detection: Hyperparameter Tuning Approach with Nested Cross-Validation," *Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi*, vol. 11, no. 1, pp. 98–110, Feb. 2026, doi: 10.25139/inform.v11i1.11221.
- [16] D. Wilimitis and C. G. Walsh, "Practical Considerations and Applied Examples of Cross-Validation for Model Development and Evaluation in Health Care: Tutorial," *JMIR AI*, vol. 2, p. e49023, Dec. 2023, doi: 10.2196/49023.
- [17] S. P. Barad, S. Kumar, and S. Mishra, "Estimation of Electronic Band Gap Energy From Material Properties Using Machine Learning," in *2024 1st International Conference on Cognitive, Green and Ubiquitous Computing (IC-CGU)*, IEEE, Mar. 2024, pp. 1–6. doi: 10.1109/IC-CGU58078.2024.10530748.
- [18] A. Demircioğlu, "Applying oversampling before cross-validation will lead to high bias in radiomics," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-62585-z.
- [19] P. Vuttipittayamongkol, E. Elyan, and A. Petrovski, "On the class overlap problem in imbalanced data classification," *Knowl. Based. Syst.*, vol. 212, p. 106631, Jan. 2021, doi: 10.1016/j.knosys.2020.106631.
- [20] M. Moradi and J. Hamidzadeh, "Handling class imbalance and overlap with a Hesitation-based instance selection method," *Knowl. Based. Syst.*, vol. 294, p. 111745, Jun. 2024, doi: 10.1016/j.knosys.2024.111745.
- [21] Q. Tao, P. Xu, M. Li, and W. Lu, "Machine learning for perovskite materials design and discovery," Dec. 01, 2021, *Nature Research*. doi: 10.1038/s41524-021-00495-8.
- [22] V. Tummalapalli, "Using SMOTE and Tomek Link Sampling Techniques to Address Imbalanced Data Challenges in the Machine Learning models."
- [23] Q. Dai, J. wei Liu, and Y. Liu, "Multi-granularity relabeled under-sampling algorithm for imbalanced data," *Appl. Soft Comput.*, vol. 124, p. 109083, Jul. 2022, doi: 10.1016/j.asoc.2022.109083.
- [24] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, Dec. 2025, doi: 10.1186/s40537-025-01313-4.
- [25] D. Wulan Yekti Rahayu *et al.*, "Performance of Machine Learning Algorithms on Imbalanced Sentiment Datasets Without Balancing Techniques," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [26] O. Rainio, J. Teuvo, and R. Klén, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, vol. 14, no. 1, Dec. 2024, doi: 10.1038/s41598-024-56706-x.
- [27] A. Tharwat, "Classification assessment methods," *Applied Computing and Informatics*, vol. 17, no. 1, pp. 168–192, Jan. 2021, doi: 10.1016/j.aci.2018.08.003.
- [28] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, vol. 16, no. 1, Dec. 2023, doi: 10.1186/s13040-023-00322-4.
- [29] M. Bhagat and B. Bakariya, "A Comprehensive Review of Cross-Validation Techniques in Machine Learning."
- [30] S. Szeghalmy and A. Fazekas, "A Comparative Study of the Use of Stratified Cross-Validation and Distribution-Balanced Stratified Cross-Validation in Imbalanced Learning," *Sensors*, vol. 23, no. 4, Feb. 2023, doi: 10.3390/s23042333.
- [31] T. Abedin, H. Xu, and S. Uddin, "The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models," *Sci. Rep.*, vol. 16, no. 1, Dec. 2026, doi: 10.1038/s41598-026-37247-x.
- [32] E. T. Anzaku, H. Wang, A. Babalola, A. Van Messum, and W. De Neve, "Re-assessing accuracy degradation: a framework for understanding DNN behavior on similar-but-non-identical test datasets," *Mach. Learn.*, vol. 114, no. 3, Mar. 2025, doi: 10.1007/s10994-024-06693-x.
- [33] L. Ferrer, O. Scharenborg, and T. Bäckström, "Good practices for evaluation of machine learning systems," Dec. 2024, [Online]. Available: <http://arxiv.org/abs/2412.03700>
- [34] M. Sivakumar, S. Parthasarathy, and T. Padmapriya, "Trade-off between training and testing ratio in machine learning for medical image processing," *PeerJ Comput. Sci.*, vol. 10, 2024, doi: 10.7717/PEERJ-CS.2245.
- [35] S. Kapse, M. Voccia, F. Viñes, and F. Illas, "Chemical bonding and electronic properties along Group 13 metal oxides," *J. Mol. Model.*, vol. 30, no. 6, Jun. 2024, doi: 10.1007/s00894-024-05957-6.
- [36] Q. Gao, R. Gao, J. Kang, and S.-H. Wei, "The Role of Ionicity in Transparent Conducting Materials," *J. Phys. Chem. Lett.*, vol. 16, no. 33, pp. 8474–8479, Aug. 2025, doi: 10.1021/acs.jpcclett.5c01925.
- [37] A. Naskar, R. Khanal, and S. Choudhury, "Role of chemistry and crystal structure on the electronic defect states in cs-based halide perovskites," *Materials*, vol. 14, no. 4, pp. 1–14, Feb. 2021, doi: 10.3390/ma14041032.