

KLASIFIKASI JENIS KACA BERDASARKAN ATRIBUT KIMIA MENGGUNAKAN METODE *K-NEAREST NEIGHBOR* (KNN)

Willy Darmadi Surya Wardhana^{1*}, Rangga Hutama Putra², Havid Danggo Pamungkas³, Anita Dewi Kusuma⁴, Arum Sari Setia Budi⁵, Dewi Oktafiani⁶

Prodi S1 Informatika, STMIK AMIKOM Surakarta

Korespodensi e-mail: ¹willy.10336@mhs.amikomsolo.ac.id,

²rangga.10471@mhs.amikomsolo.ac.id, ³havid.10424@mhs.amikomsolo.ac.id,

⁴anita.10437@mhs.amikomsolo.ac.id, ⁵arum.10433@mhs.amikomsolo.ac.id,

⁶dewioktafiani@dosen.amikomsolo.ac.id

Abstrak

Kaca merupakan material penting dalam industri manufaktur, konstruksi, dan otomotif. Penentuan jenis kaca yang tepat diperlukan untuk menjaga kualitas, keamanan, dan efisiensi produksi. Namun, proses identifikasi secara manual sering membutuhkan keahlian khusus dan rentan terhadap kesalahan. Oleh karena itu, diperlukan sistem klasifikasi berbasis data untuk membantu proses pengenalan jenis kaca. Penelitian ini menerapkan metode *K-Nearest Neighbors* (KNN) untuk mengklasifikasikan tujuh jenis kaca berdasarkan sembilan atribut kimia. Evaluasi model menggunakan metrik akurasi, presisi, recall, *F1-score*, serta confusion matrix. Hasil penelitian menunjukkan bahwa nilai *K* terbaik adalah *K*=2 dengan akurasi 72%. Jenis kaca kategori 5 memperoleh performa terbaik dengan precision, recall, dan *F1-score* sebesar 1.00. Sementara itu, kategori 2 dan 3 menunjukkan performa lebih rendah karena kemiripan karakteristik antar kelas. Temuan ini menunjukkan bahwa KNN cukup efektif untuk klasifikasi jenis kaca, meskipun masih memerlukan optimasi lebih lanjut pada kelas tertentu.

Kata kunci: kaca, klasifikasi, evaluasi model, *k-nearest neighbors*

Abstract

Glass is an essential material widely used in manufacturing, construction, and automotive industries. Determining the correct type of glass is crucial to ensuring product quality, safety, and production efficiency. However, manual identification often requires specialized expertise and is prone to human error. Therefore, a data-driven classification system is needed to support the glass identification process. This study applies the *K-Nearest Neighbors* (KNN) method to classify seven types of glass based on nine chemical attributes. The model is evaluated using accuracy, precision, recall, *F1-score*, and a confusion matrix. The results show that the optimal value of *K* is 2, achieving an accuracy of 72%. Glass type 5 demonstrates the best performance, with precision, recall, and *F1-score* all reaching 1.00. Meanwhile, types 2 and 3 show lower performance due to similarities in their chemical characteristics. These findings indicate that KNN is sufficiently effective for glass classification, although further optimization is still needed for certain classes.

Keywords: glass, classification, model evaluation, *k-nearest neighbors*

PENDAHULUAN

Kaca merupakan salah satu material penting yang banyak digunakan dalam berbagai bidang seperti konstruksi, otomotif, elektronik, dan industri rumah tangga. Setiap jenis kaca memiliki komposisi kimia dan karakteristik fisik yang berbeda sehingga pemilihannya harus disesuaikan dengan kebutuhan penggunaan[1]. Dalam proses industri, kesalahan dalam mengidentifikasi jenis kaca dapat menyebabkan penurunan kualitas produk, potensi kegagalan material, serta inefisiensi proses produksi. Oleh karena itu, identifikasi jenis kaca yang akurat dan konsisten menjadi kebutuhan penting dalam industri modern[2].

Proses identifikasi kaca secara manual umumnya dilakukan melalui pengukuran komposisi kimia atau berdasarkan pengalaman teknisi. Namun, metode manual memiliki beberapa keterbatasan seperti subjektivitas, membutuhkan waktu yang lama, serta berisiko terjadi kesalahan terutama pada jenis kaca yang memiliki karakteristik kimia yang mirip. Hal ini menimbulkan kebutuhan akan sistem klasifikasi berbasis data untuk membantu proses identifikasi secara lebih cepat, objektif, dan akurat[3].

Machine learning menyediakan pendekatan analitis yang mampu mempelajari pola hubungan antar atribut kimia pada berbagai jenis kaca. Berbagai metode klasifikasi seperti *Support Vector Machine* (SVM), *Decision Tree* (DT), *Random Forest*, dan *Naïve Bayes* telah digunakan dalam berbagai penelitian untuk menangani permasalahan serupa. Namun, setiap metode memiliki karakteristik, kelebihan, dan keterbatasan masing-masing[4]. Metode berbasis pohon keputusan cenderung menghasilkan model yang mudah *overfitting*, sedangkan metode probabilistik seperti *Naïve Bayes* mengasumsikan independensi antar fitur yang tidak selalu sesuai dengan kondisi data kimia kaca. Sementara itu, metode seperti SVM membutuhkan proses pelatihan kompleks dan sensitif terhadap pemilihan parameter kernel.

Dalam konteks dataset kaca yang memiliki jumlah data relatif kecil, bersifat numerik, serta menunjukkan kemiripan antar kelas, metode *K-Nearest Neighbors* (KNN) menawarkan pendekatan yang sederhana namun efektif. KNN merupakan algoritma non-parametrik yang mengklasifikasikan suatu sampel berdasarkan kedekatan nilai-nilai fitur dengan sampel lain dalam dataset[5]. Metode ini tidak memerlukan pelatihan model yang rumit dan terbukti memberikan performa yang stabil pada dataset berukuran kecil dan menengah. Beberapa penelitian terdahulu juga menunjukkan bahwa KNN dapat memberikan akurasi tinggi pada kasus klasifikasi berbasis data kimia[6].

Berdasarkan informasi tersebut, studi ini bertujuan untuk menggunakan metode *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan tipe kaca dengan memanfaatkan sembilan atribut kimia yang ada dalam dataset *UCI Machine Learning Repository*. Metode ini dipilih karena mampu mendeteksi pola non-linear dan mengelompokkan data sesuai dengan kesamaan nilai antaratribut tanpa harus berpegang pada asumsi distribusi tertentu [7]. Evaluasi kinerja model dilaksanakan dengan memanfaatkan ukuran akurasi, *presisi*, *recall*, *F1-score*, dan juga *confusion matrix* untuk mengevaluasi sejauh mana KNN dapat mengenali ciri-ciri masing-masing jenis kaca. Penelitian ini diharapkan mampu memberikan sumbangan dalam bidang ilmiah terkait penerapan algoritma KNN untuk pengelompokan material berdasarkan sifat kimia, serta memberikan manfaat praktis dalam pengembangan sistem identifikasi kaca yang otomatis dan efektif.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk menganalisis performa algoritma *K-Nearest Neighbors* (KNN) dalam mengklasifikasikan jenis kaca berdasarkan atribut kimia. Tahapan penelitian dirancang secara sistematis agar proses

analisis berjalan terstruktur dan dapat direplikasi. Adapun alur penelitian meliputi pengumpulan data, pra-pemrosesan data, analisis eksploratif, pemodelan, dan evaluasi performa.

Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berasal dari Kaggle pada tautan <https://www.kaggle.com/datasets/uciml/glass>. Dataset ini terdiri dari 214 sampel kaca dengan sembilan atribut kimia, yaitu: *Refractive Index* (RI), Sodium (Na), Magnesium (Mg), Aluminum (Al), Silicon (Si), Potassium (K), Calcium (Ca), Barium (Ba), Iron (Fe), serta label target berupa jenis kaca (*Type of Glass*). Terdapat tujuh jenis kaca dalam dataset, namun tipe ke-4 (*vehicle windows non-float processed*) tidak tersedia. Dataset digunakan sebagaimana adanya untuk proses klasifikasi[8].

Pra-pemrosesan Data (*Pre-processing*)

Pra-pemrosesan dilakukan untuk memastikan data siap digunakan dalam tahap pemodelan. Tahapan pra-pemrosesan meliputi:

- a) Pemeriksaan data duplikat dan penghapusan data yang berulang.
- b) Pemeriksaan data kosong (*missing values*); dataset tidak mengandung nilai kosong sehingga tidak diperlukan imputasi data.
- c) Normalisasi fitur dilakukan apabila diperlukan untuk menyetarakan skala antar variabel numerik sehingga perhitungan jarak pada KNN menjadi lebih stabil.

Pra-pemrosesan ini bertujuan meningkatkan kualitas data serta mengurangi bias yang dapat memengaruhi performa model.

Analisis Eksploratif Data (*Exploratory Data Analysis*)

Tahap ini dilakukan untuk memahami karakteristik distribusi data, hubungan antar variabel, serta sebaran jenis kaca. Visualisasi dilakukan terhadap nilai tiap atribut untuk mengetahui apakah terdapat cluster alami atau kemiripan antar jenis kaca. Analisis ini penting untuk memberikan gambaran awal terkait potensi kesulitan model dalam membedakan kelas yang memiliki fitur sangat mirip[9].

Pembagian Data (*Data Splitting*)

Dataset dibagi menjadi data latih (*training*) sebesar 80% dan data uji (*testing*) sebesar 20%. Pembagian ini menggunakan teknik *stratified split* untuk menjaga proporsi kelas tetap seimbang pada kedua subset data[10]. Data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi generalisasi model.

Pemodelan Menggunakan *K-Nearest Neighbors*

Tahap ini merupakan inti dari penelitian, yaitu penerapan algoritma KNN untuk klasifikasi jenis kaca. Proses pemodelan meliputi:

1. Menentukan nilai K (jumlah tetangga terdekat) yang akan diuji.
2. Melakukan proses pencarian nilai K optimal dengan mencoba beberapa nilai K (misalnya 2, 3, 4, 5).
3. Menggunakan jarak Euclidean sebagai ukuran kedekatan antar sampel.
4. Menghasilkan model final berdasarkan nilai K dengan akurasi terbaik.

Pemilihan K didasarkan pada performa model pada data uji. Nilai K yang terlalu kecil dapat menyebabkan model sensitif terhadap *noise*, sedangkan nilai K yang terlalu besar dapat membuat model menjadi terlalu bias.

Evaluasi Model

Evaluasi model dilakukan pada data uji menggunakan metrik:

- Akurasi: mengukur *persentase* prediksi yang benar.
- $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- Presisi (*precision*): tingkat ketepatan prediksi positif untuk setiap kelas.
- $$Precision = \frac{TP}{TP + FP} \quad (2)$$

- Recall: kemampuan model mendeteksi kelas sebenarnya.
- $$Recall = \frac{TP}{TP + FN} \quad (3)$$

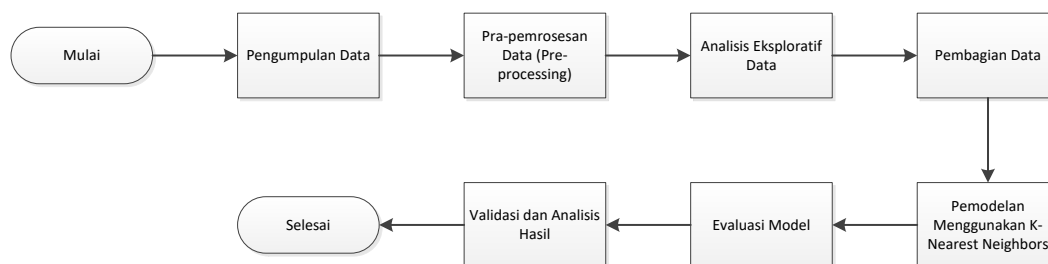
- F1-score: rata-rata harmonik antara presisi dan *recall*.
- $$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (4)$$

- *Confusion matrix*: memberikan gambaran detail mengenai distribusi prediksi benar dan salah antar kelas.

Evaluasi ini digunakan untuk melihat performa keseluruhan model serta menilai kemampuan model dalam mengklasifikasikan setiap jenis kaca secara spesifik.

Validasi dan Analisis Hasil

Hasil evaluasi dianalisis untuk mengidentifikasi kelemahan model, terutama pada kelas yang memiliki kemiripan fitur sehingga rawan salah klasifikasi. Analisis ini menjadi dasar untuk memberikan rekomendasi pengembangan metode berikutnya, seperti balancing data, pemilihan fitur, atau penggunaan algoritma lain untuk perbandingan performa[11].



Gambar 1. Alur Metode Penelitian

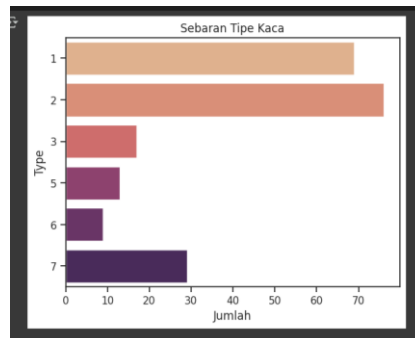
HASIL DAN PEMBAHASAN

Deskripsi Umum Dataset

Dataset kaca terdiri dari 214 sampel dengan sembilan atribut kimia serta enam jenis kaca yang tersedia (tipe 1, 2, 3, 5, 6, dan 7). Distribusi kelas tidak merata, di mana beberapa jenis kaca memiliki jumlah sampel yang jauh lebih sedikit, seperti tipe 3, 5, dan 6. Ketidakseimbangan ini berpotensi memengaruhi performa model, terutama pada kelas minoritas yang lebih rentan salah diklasifikasikan. Analisis awal terhadap nilai fitur menunjukkan bahwa beberapa jenis kaca memiliki kemiripan komposisi kimia, khususnya pada variabel Na, Mg, dan Ca, sehingga model berbasis jarak seperti KNN berpotensi mengalami kesulitan dalam membedakan kelas yang berdekatan secara numerik[12].

Hasil Pra-pemrosesan dan Analisis Data

Pra-pemrosesan menunjukkan bahwa tidak terdapat nilai kosong pada dataset, namun terdapat beberapa data duplikat yang perlu dihapus untuk menjaga keaslian distribusi data. Analisis eksploratif memperlihatkan bahwa atribut RI, Na, dan Ca memiliki pengaruh signifikan dalam membedakan beberapa kelas, sementara atribut seperti Fe memiliki nilai yang sangat rendah dan tidak terlalu bervariasi[13]. Visualisasi distribusi data mengindikasikan bahwa sebagian kelas memiliki overlap yang besar, sehingga menantang bagi KNN dalam memisahkan batas antar kelas secara jelas.



Gambar 2. Visualisasi Sebaran Data

Penentuan Nilai K Optimal

Eksperimen nilai K dilakukan untuk mengetahui jumlah tetangga terdekat yang memberikan performa terbaik. Nilai K yang diuji meliputi 2, 3, 4, dan 5. Hasil evaluasi menunjukkan bahwa:

1. $K = 2$ memberikan akurasi tertinggi, yaitu 72%.
2. Nilai K yang lebih besar cenderung menurunkan akurasi karena model menjadi terlalu general, sehingga sampel dari kelas minoritas sering terserap ke kelas mayoritas.

Tabel 1. Akurasi berbagai nilai K

Nilai K	Akurasi
2	72%
3	67%
4	58%
5	67%

Hal ini sejalan dengan karakteristik dataset yang kecil dan memiliki kelas tidak seimbang, sehingga nilai K yang rendah lebih mampu mempertahankan informasi lokal dari data.

Hasil Confusion Matrix

Dari model yang dibuat, didapatkan confusion matrix pada tabel dibawah ini:

Tabel 2. Confusion Matrix

Actual\Predicted	Type 1	Type 2	Type 3	Type 5	Type 6	Type 7
Type 1	10	0	0	0	0	0
Type 2	7	8	0	0	0	0
Type 3	2	0	1	0	0	0
Type 5	0	0	0	3	0	0
Type 6	0	1	0	0	2	0
Type 7	1	1	0	0	0	7

Confusion matrix menunjukkan bahwa tipe 1, 5, dan 7 memiliki tingkat klasifikasi yang sangat baik. Tipe 2 dan 3 memiliki tingkat kesalahan tinggi dan sering diprediksi sebagai tipe 1. Tipe 6 memiliki hasil cukup baik meskipun terdapat beberapa kesalahan prediksi.

Kesalahan klasifikasi terutama terjadi pada kelas yang secara statistik memiliki kemiripan atribut seperti Na, Mg, dan Ca. Misalnya, tipe 2 dan 3 memiliki rentang nilai yang berdekatan pada beberapa fitur sehingga jarak *Euclidean* sulit membedakan sampel secara jelas[14].

Classification Report

Pada *classification report* terdapat *precision*, *recall*, *f1-score*, dan *support* untuk masing-masing tipe. Berikut tabel hasil *classification report*.

Tabel 3. Classification Report

Type	Precision	Recall	F1-Score	Support
1	0.50	1.00	0.67	10
2	0.80	0.53	0.64	15
3	1.00	0.33	0.50	3
5	1.00	1.00	1.00	3
6	1.00	0.67	0.80	3
7	1.00	0.78	0.88	9

Hasil evaluasi berdasarkan *precision*, *recall*, dan *F1-score* menunjukkan tipe 5 memiliki performa sempurna. Tipe 7 dan tipe 6 memiliki *precision* tinggi namun *recall* lebih rendah, menunjukkan bahwa hanya sebagian sampel yang dapat dikenali oleh model. Tipe 1, 2, dan 3 memiliki performa tidak seimbang, terutama tipe 3 yang memiliki *recall* hanya 0.33.

Performa buruk pada tipe 3 dan tipe 2 dapat disebabkan beberapa hal seperti jumlah sampel yang sedikit sehingga informasi pembeda menjadi terbatas, overlap fitur yang signifikan dengan tipe 1 yang menyebabkan model lebih sering menarik sampel tersebut ke kelas yang lebih besar, algoritma KNN yang sensitif terhadap distribusi dan skala fitur sehingga perbedaan kecil dalam nilai kimia dapat menyebabkan kesalahan klasifikasi.

Analisis dan Interpretasi Hasil

Secara umum, hasil menunjukkan bahwa KNN cukup efektif digunakan untuk jenis kaca tertentu, namun tidak secara merata pada semua kelas. Kinerja terbaik diperoleh pada kelas yang memiliki distribusi fitur yang unik dan tidak banyak *overlap*, seperti tipe 5. Sebaliknya, kelas dengan karakteristik kimia yang sangat mirip cenderung sulit dibedakan[15].

Beberapa poin penting dari hasil penelitian:

Sensitivitas terhadap Kedekatan Nilai Fitur

KNN menggunakan jarak *Euclidean* sehingga kelas yang memiliki jarak antar fitur yang sangat kecil akan sulit dibedakan. Hal ini terlihat jelas pada tipe 1, 2, dan 3.

Pengaruh Ketidakseimbangan Kelas

Kelas minoritas cenderung kalah dominan terhadap kelas mayoritas, sehingga model sering memproyeksikan data minoritas ke kelas yang lebih besar.

Potensi Peningkatan Model

Model berpotensi ditingkatkan melalui normalisasi fitur yang lebih optimal, teknik penyeimbangan data seperti SMOTE, pemilihan fitur (*feature selection*), atau penggunaan metode lain seperti *Random Forest* atau SVM untuk perbandingan.

Ringkasan Hasil

Secara keseluruhan, KNN mampu mengklasifikasikan beberapa jenis kaca dengan baik, khususnya tipe yang memiliki perbedaan nilai kimia yang jelas. Namun, model mengalami keterbatasan pada jenis dengan karakteristik mirip dan jumlah sampel sedikit. Akurasi terbaik sebesar 72% menunjukkan bahwa KNN dapat menjadi solusi awal, tetapi penelitian lanjutan diperlukan untuk meningkatkan performa model secara keseluruhan.

KESIMPULAN

Penelitian ini telah sukses menerapkan metode *K-Nearest Neighbor* (KNN) untuk mengkategorikan tipe kaca berdasarkan sembilan atribut kimia dari *dataset UCI Machine Learning Repository*. Hasil percobaan menunjukkan bahwa metode KNN dapat dengan baik mengidentifikasi karakteristik masing-masing jenis kaca, di mana nilai akurasi, *presisi*, *recall*, dan *F1-score* yang diperoleh mencerminkan performa model yang cukup konsisten dalam mengelompokkan data. Dari evaluasi yang dilakukan, nilai K yang paling optimal ditemukan pada $K = 2$, dengan tingkat akurasi tertinggi di antara variasi K yang diuji. Penemuan ini menunjukkan bahwa algoritma KNN bisa menjadi metode yang efektif dalam pengklasifikasian bahan berdasarkan komposisi kimiawi. Penelitian ini memberikan sumbangan ilmiah dalam penerapan algoritma pembelajaran mesin untuk analisis material dan dapat menjadi dasar untuk mengembangkan model klasifikasi yang lebih rumit di masa depan, seperti dengan menambahkan proses pemilihan fitur atau membandingkannya dengan algoritma lain yang lebih gesit.

DAFTAR PUSTAKA

- [1]R. Marin and D. Alvarez, "Glass Materials: Structure, Properties, and Applications," *J. Mater. Sci. Eng.*, vol. 45, no. 3, pp. 112–130, 2021, doi: 10.1016/j.mateng.2021.03.004.
- [2]A. Hartono and T. Wijaya, "Perkembangan Material Kaca dalam Industri Modern," *J. Teknol. Mater.*, vol. 14, no. 2, pp. 101–112, 2022.
- [3]D. H. Putra and M. Sari, "Analisis Kesalahan Identifikasi Material Berbasis Pengamatan Manual," *J. Rekayasa Ind.*, vol. 19, no. 1, pp. 33–41, 2023.
- [4]A. Rahman and F. Kurniawan, "Penerapan metode machine learning untuk klasifikasi material berbasis komposisi kimia," *J. Inform. dan Sains*, vol. 8, no. 4, pp. 299–308, 2021.
- [5]L. Zhao and Y. Chen, "Enhanced KNN Classification Performance on Small Numerical Datasets," *Int. J. Data Sci. Anal.*, vol. 18, no. 2, pp. 89–104, 2022, doi: 10.1007/s41060-022-00312-7.
- [6]R. Nugroho and D. Lestari, "Evaluasi Metode KNN pada Dataset Numerik Berukuran Kecil," *J. Teknol. Inf. Indones.*, vol. 7, no. 3, pp. 221–230, 2022.
- [7]B. Santoso and G. Pratama, "Evaluasi model klasifikasi menggunakan akurasi, presisi, recall, dan F1-score pada data material," *J. Sains Komputasi*, vol. 12, no. 1, pp. 44–56, 2024.
- [8]R. Patel and A. Singh, "Glass Identification Dataset: A modern analysis approach," *J. Data Resour.*, vol. 5, no. 2, pp. 77–85, 2023.
- [9]A. Putri and D. Yuniar, "Analisis EDA pada dataset numerik untuk identifikasi pola awal," *J. Ilmu Komputasi Indones.*, vol. 9, no. 1, pp. 13–22, 2023.
- [10] F. Susanto and S. Rahmawati, "Evaluasi teknik stratified split untuk menjaga sebaran

- kelas pada dataset tidak seimbang,” *J. Data Sains Indones.*, vol. 4, no. 3, pp. 101–110, 2022.
- [11] A. Widodo and P. Lestari, “Analisis kesalahan klasifikasi pada algoritma supervised learning,” *J. Sains Komput. Terap.*, vol. 7, no. 1, pp. 45–56, 2023.
- [12] R. Sunarto and B. Wijaya, “Tantangan algoritma berbasis jarak pada dataset dengan fitur mirip,” *J. Teknol. Inf. Indones.*, vol. 11, no. 3, pp. 77–88, 2022.
- [13] D. Kusuma and F. Ardiansyah, “Pengaruh variabilitas fitur terhadap kemampuan pemisahan kelas pada algoritma KNN,” *J. Algoritm. dan Komputasi*, vol. 12, no. 1, pp. 31–44, 2023.
- [14] K. Sari and D. Pratama, “Evaluasi kesalahan klasifikasi pada algoritma KNN berbasis jarak Euclidean,” *J. Teknol. Inf. Nusantara*, vol. 4, no. 2, pp. 76–85, 2022.
- [15] A. Rahman and T. Setiani, “Analisis performa KNN berdasarkan variasi distribusi fitur pada data klasifikasi material,” *J. Ilmu Komput. dan Teknol.*, vol. 9, no. 4, pp. 122–134, 2022.