

Analisis Pengaruh Penanganan Imbalanced Dataset Terhadap Kinerja Model Klasifikasi Kualitas Peta Konsep Menggunakan Algoritma Support Vector Machine

Kotot Tri Hartanto^a, Purnawansyah^b, Siti Rahmah Jabir^c

Universitas Muslim Indonesia, Makassar, Indonesia

^a13020230366@umi.ac.id; ^bpurnawansyah@umi.ac.id; ^crahmah@umi.ac.id

Received: 24-02-2026 | Revised: 27-02-2026 | Accepted: 08-03-2026 | Published: 29-03-2026

Abstrak

Penilaian kualitas peta konsep sering digunakan untuk melihat pemahaman mahasiswa secara lebih terstruktur, namun proses penilaiannya masih banyak dilakukan secara manual sehingga memerlukan waktu dan berpotensi tidak konsisten. Tantangan lain yang muncul adalah ketidakseimbangan jumlah data antar kelas kualitas, yang dapat membuat model klasifikasi cenderung bias pada kelas mayoritas. Penelitian ini bertujuan menguji pengaruh SMOTE ENN terhadap kinerja *Support Vector Machine* dalam mengklasifikasikan kualitas proposisi peta konsep. Data yang digunakan berjumlah 691 proposisi dengan empat kelas kualitas, di mana kelas 3 mendominasi sebesar 64,3 persen. Teks diproses melalui tahap pra pemrosesan, kemudian direpresentasikan menggunakan TF IDF dengan ukuran kosakata 321 fitur. Data dibagi menjadi pelatihan dan pengujian dengan rasio 80 banding 20. Eksperimen dilakukan pada dua skenario, yaitu SVM tanpa penyeimbangan sebagai *baseline*, serta SVM dengan SMOTE ENN yang diterapkan pada data pelatihan sehingga ukuran data pelatihan meningkat dari 552 menjadi 1309 sampel dan distribusi kelas menjadi lebih seimbang. Hasil pengujian menunjukkan bahwa SMOTE ENN meningkatkan *recall* rata-rata antar kelas dari 0,6479 menjadi 0,7149 dan *F1 Score* rata-rata antar kelas dari 0,6806 menjadi 0,7297, serta menurunkan MAE dari 0,2230 menjadi 0,1871. Akurasi mengalami penurunan kecil dari 0,8633 menjadi 0,8561. Namun, penurunan ini disertai peningkatan *recall* dan *F1* rata-rata antar kelas, yang berarti model lebih adil dan lebih baik dalam mendeteksi kelas-kelas yang sebelumnya jarang. Temuan ini mengindikasikan bahwa SMOTE ENN efektif untuk meningkatkan pemerataan performa klasifikasi pada dataset peta konsep yang tidak seimbang.

Kata kunci: peta konsep, klasifikasi kualitas, SMOTE ENN, TF-IDF

Pendahuluan

Peta konsep banyak digunakan dalam pembelajaran karena mampu menggambarkan cara berpikir, keterkaitan antar konsep, serta kedalaman pemahaman peserta didik terhadap suatu materi. Dalam praktiknya, peta konsep tidak hanya menjadi media belajar, tetapi juga menjadi instrumen penilaian untuk melihat kualitas pemahaman secara lebih terstruktur [1], [2]. Tantangannya, proses penilaian peta konsep secara manual membutuhkan waktu, cenderung melelahkan, dan berpotensi menimbulkan perbedaan penilaian antar penilai. Kondisi ini mendorong kebutuhan akan pendekatan otomatis yang lebih konsisten dan efisien untuk membantu proses evaluasi.

Upaya otomatisasi penilaian peta konsep umumnya diarahkan pada tugas klasifikasi kualitas, yaitu mengelompokkan peta konsep ke dalam beberapa tingkat kualitas seperti baik, cukup, kurang, dan sangat kurang. Dengan adanya model klasifikasi, proses penilaian diharapkan dapat dilakukan lebih cepat sekaligus tetap mempertahankan ketelitian. Namun, penerapan klasifikasi kualitas peta konsep sering berhadapan dengan persoalan yang cukup mendasar, yaitu ketidakseimbangan jumlah data antar kelas. Pada banyak dataset, kelas mayoritas seperti kategori baik memiliki jumlah sampel yang jauh lebih besar dibanding kelas minoritas seperti kurang atau sangat kurang [3], [4], [5].

Ketidakeimbangan data ini dapat membuat model belajar secara tidak adil. Model cenderung menebak kelas mayoritas karena hal tersebut tampak paling menguntungkan dari sisi jumlah data, sementara kemampuan mengenali kelas minoritas menjadi rendah [6], [7]. Dalam konteks evaluasi pendidikan, kelemahan ini bukan hanya masalah teknis, karena kesalahan pada kelas minoritas justru bisa berdampak besar. Peta konsep dengan kualitas rendah yang tidak terdeteksi, atau peta konsep yang seharusnya dinilai tinggi namun tertukar kelas, dapat mengurangi nilai guna sistem penilaian otomatis. Akibatnya, keluaran model menjadi kurang representatif dan tidak mencerminkan kondisi sebenarnya secara menyeluruh.

Berdasarkan permasalahan di atas, penelitian ini menempatkan penanganan ketidakseimbangan data sebagai fokus utama untuk meningkatkan keandalan klasifikasi. Salah satu pendekatan yang banyak digunakan adalah *resampling*, yaitu menyesuaikan komposisi data melalui penambahan sampel pada kelas minoritas atau pengurangan sampel pada kelas mayoritas. Pendekatan gabungan dinilai menarik karena tidak hanya menambah representasi kelas minoritas, tetapi juga berusaha membersihkan data yang berpotensi mengganggu pembelajaran model. Di antara teknik *hybrid* yang relevan, SMOTE ENN menawarkan mekanisme pembentukan sampel sintetik untuk kelas minoritas sekaligus penyaringan sampel ambigu melalui prinsip tetangga terdekat.

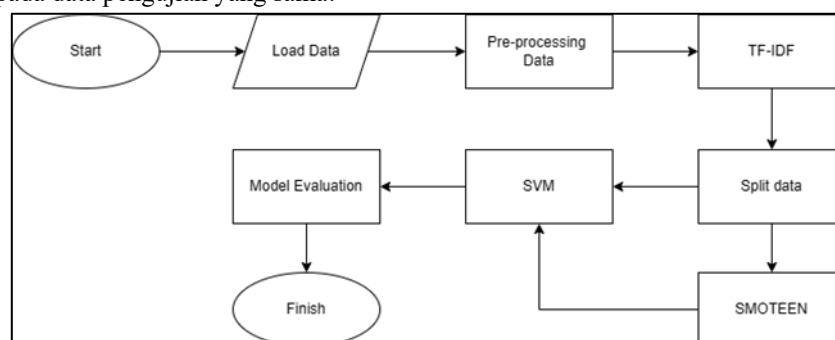
Penelitian ini mengkaji pengaruh penerapan SMOTE ENN terhadap kinerja model klasifikasi kualitas peta konsep berbasis teks [8], [9]. Dataset peta konsep yang telah diberi label kualitas terlebih dahulu melalui tahap pra pemrosesan seperti pembersihan data, normalisasi huruf, tokenisasi, penghapusan kata umum (stopword removal), serta stemming untuk menyederhanakan variasi kata ke bentuk dasarnya. Representasi fitur dibangun menggunakan TF IDF untuk menangkap tingkat kepentingan kata di dalam dokumen [10], [11]. Setelah itu, data dibagi menjadi data pelatihan dan pengujian dengan rasio 80 banding 20, lalu SMOTE ENN diterapkan pada data pelatihan sebelum model dibangun.

Algoritma yang digunakan adalah *Support Vector Machine* karena dikenal kuat untuk data berdimensi tinggi, termasuk vektor TF IDF, serta mampu membentuk batas keputusan yang baik. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, *recall*, *F1 Score*, dan juga MAE [12]. Penggunaan MAE menjadi relevan karena label kualitas memiliki sifat bertingkat sehingga selisih prediksi dapat memberikan gambaran seberapa jauh kesalahan model dari tingkat kualitas yang semestinya. Analisis dilakukan dengan membandingkan performa sebelum dan sesudah penanganan ketidakseimbangan, termasuk melalui confusion matrix untuk melihat perubahan pola kesalahan antar kelas.

Melalui pendekatan ini, penelitian diharapkan mampu menunjukkan bahwa penanganan ketidakseimbangan data dapat meningkatkan kemampuan model dalam mengenali kelas minoritas tanpa mengorbankan performa secara tidak perlu. Pada akhirnya, hasil penelitian ini diharapkan memberi dasar yang lebih kuat untuk pengembangan sistem penilaian otomatis peta konsep yang lebih adil, reliabel, dan sesuai kebutuhan evaluasi pendidikan, sekaligus menegaskan pentingnya pemilihan metrik evaluasi yang tepat pada dataset yang tidak seimbang.

Metode

Penelitian ini dirancang sebagai eksperimen komparatif untuk menilai pengaruh penanganan data tidak seimbang terhadap kinerja klasifikasi kualitas peta konsep berbasis teks. Eksperimen dilakukan melalui dua skenario. Skenario pertama menggunakan data pelatihan apa adanya tanpa penyeimbangan kelas. Skenario kedua menerapkan SMOTE ENN pada data pelatihan, lalu model dilatih kembali. Hasil kedua skenario dibandingkan pada data pengujian yang sama.



Gambar 1. Desain penelitian

A. Data penelitian

Data yang digunakan dalam penelitian ini diperoleh dari kumpulan proposisi mengenai topik basis data relasional yang telah dinilai oleh seorang pakar (expert rated). Dataset ini tersedia dalam format Excel (.xlsx) dengan nama file Dataset_Basisdata_Relasional_RATED_1_expert.xlsx. Dataset berisi 691 baris

data dengan empat atribut utama, yaitu: UID, No, Proposisi, dan *Score*. Atribut Proposisi berisi teks pernyataan atau relasi konsep yang akan dijadikan fitur dalam klasifikasi, sedangkan atribut *Score* merupakan label yang menunjukkan tingkat kualitas atau ketepatan proposisi berdasarkan penilaian pakar, dengan nilai berupa angka 0, 1, 2, atau 3. UID menunjukkan identitas unik untuk setiap peta konsep, dan No menunjukkan urutan proposisi dalam masing-masing UID [1], [13]. Dalam penelitian ini, atribut Proposisi akan digunakan sebagai fitur input, sedangkan atribut *Score* akan menjadi target klasifikasi.

Tabel 1. Distribusi kelas *score*

No	Label	Jumlah	Persentase (%)
1	0	38	5.5
2	1	77	11.14
3	2	132	19.1
4	3	444	64.25

Sebaran ini memperlihatkan *score* 3, sehingga diperlukan strategi penanganan model tidak hanya kuat pada kelas mayoritas.

B. Pra pemrosesan teks

Karena data berbentuk teks, tahap pra pemrosesan dilakukan agar representasi fitur lebih bersih dan konsisten. Tahapan yang digunakan meliputi pembersihan karakter tidak relevan (misalnya URL dan simbol), penyeragaman huruf menjadi huruf kecil, tokenisasi berbasis spasi, penghapusan kata umum yang kurang informatif menggunakan daftar stopword Bahasa Indonesia dari pustaka Sastrawi, serta stemming (Sastrawi Stemmer) untuk menyederhanakan variasi kata ke bentuk dasarnya. Selain itu, peneliti dapat menambahkan stopwords kustom yang sangat umum pada konteks topik namun kurang informatif. Seluruh tahapan pra pemrosesan diterapkan konsisten pada data pelatihan dan data pengujian.

C. Ekstraksi fitur menggunakan TF-IDF

Teks yang sudah diproses diubah menjadi vektor numerik menggunakan TF IDF. Pembobotan ini memberi bobot lebih tinggi pada kata yang sering muncul dalam satu dokumen tetapi jarang muncul pada dokumen lain, sehingga lebih informatif untuk membedakan kelas [14], [15], [16], [17]. Rumus dasar yang digunakan adalah sebagai berikut:

1. Rumus *Term Frequency*

$$TF(t, d) = \frac{f(t, d)}{\sum_k f(k, d)} \quad (1)$$

Dimana:

$f(t, d)$ = frekuensi kemunculan term t dalam dokumen d .

$\sum_k f(k, d)$ = total seluruh term pada dokumen d

2. Rumus *Inverse Document Frequency* (IDF)

Mengurangi bobot kata yang sering muncul dalam banyak dokumen.

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

Dimana:

N = jumlah seluruh dokumen dalam korpus

$DF(t)$ = jumlah dokumen yang mengandung term t

3. Rumus TF-IDF

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

D. Pembagian data latih dan data uji

Data dibagi menjadi data pelatihan dan data pengujian dengan rasio 80 banding 20. Dengan total 691 data, pembagian ini menghasilkan sekitar 553 data untuk pelatihan dan 138 data untuk pengujian. Pembagian dilakukan secara stratified agar proporsi setiap kelas pada data pengujian tetap mendekati kondisi asli. Penerapan SMOTE ENN hanya dilakukan pada data pelatihan untuk mencegah kebocoran informasi dari data pengujian.

E. Penanganan data tidak seimbang dengan SMOTE ENN

Untuk mengatasi ketidakseimbangan kelas, digunakan metode *hybrid* SMOTE ENN dan diterapkan pada data pelatihan. SMOTE ENN merupakan pendekatan yang menggabungkan dua metode utama dalam penanganan data tidak seimbang, yaitu oversampling melalui Synthetic Minority Over-sampling Technique (SMOTE) dan undersampling menggunakan Edited Nearest Neighbor (ENN). SMOTE bekerja dengan menghasilkan data sintesis baru pada kelas minoritas melalui interpolasi antara sampel- sampel minoritas terdekat [18], [19]. Di sisi lain, ENN berfungsi sebagai teknik pembersihan data (data cleaning) dengan menghapus sampel-sampel yang tidak sesuai dengan mayoritas tetangganya berdasarkan aturan tetangga terdekat. Kombinasi dari kedua metode ini memungkinkan distribusi kelas menjadi lebih seimbang dan noise dalam data dapat dikurangi [9], [20].

F. Pembangunan model klasifikasi menggunakan *Support Vector Machine*

Model klasifikasi dibangun menggunakan *Support Vector Machine* karena efektif untuk fitur berdimensi tinggi seperti TF IDF. Model dilatih pada dua skenario. Skenario pertama memakai data pelatihan asli. Skenario kedua memakai data pelatihan yang sudah diproses SMOTE ENN. Selain perbedaan itu, pengaturan pelatihan dibuat konsisten agar perbandingan tetap adil.

G. Evaluasi kinerja model

Evaluasi dilakukan pada data pengujian menggunakan beberapa metrik agar performa keseluruhan dan performa per kelas dapat terlihat jelas, terutama pada kelas minoritas.

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1 = \frac{2 \times Precision \times recall}{Precision + Recall} \tag{7}$$

Selain itu, digunakan MAE karena label *Score* bersifat bertingkat, sehingga jarak kesalahan prediksi juga penting untuk dinilai.

$$MAE = \frac{1}{n} \sum_{t=1}^n |\gamma_t - \hat{\gamma}_t| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\gamma_t - \hat{\gamma}_t)^2} \tag{9}$$

Confusion matrix juga digunakan untuk melihat pola kesalahan antar kelas, misalnya kelas mana yang paling sering tertukar dan apakah pengenalan kelas minoritas membaik setelah penyeimbangan.

H. Lingkungan implementasi

dSeluruh proses, mulai dari pra pemrosesan, ekstraksi fitur TF IDF, penerapan SMOTE ENN, pelatihan SVM, hingga evaluasi metrik, diimplementasikan menggunakan Python dan pustaka pembelajaran mesin yang relevan.

Hasil dan Diskusi

Bagian ini menyajikan hasil eksperimen klasifikasi kualitas proposisi peta konsep menggunakan SVM pada dua skenario, yaitu *baseline* tanpa penyeimbangan dan skenario dengan SMOTE ENN. Pembahasan difokuskan pada perubahan performa antar kelas, khususnya kelas minoritas, serta trade off yang muncul setelah data pelatihan dibuat lebih seimbang.

A. Karakteristik data dan fitur

Dataset berisi 691 proposisi dengan empat kelas kualitas. Sebaran kelas sangat timpang, karena kelas 3 mendominasi dapat dilihat persebaran datanya pada Table 1.

Setelah pra pemrosesan, rata rata panjang proposisi adalah 5,5 kata. Ekstraksi fitur menggunakan TF IDF menghasilkan kosakata 321 kata unik dengan dimensi fitur 321 dan sparsity 97,26 persen. Ini menunjukkan vektor fitur sangat jarang, tetapi masih wajar untuk pendekatan TF IDF pada teks pendek.

B. Hasil skenario *baseline* tanpa penyeimbangan

Pada *baseline*, model SVM dilatih menggunakan data pelatihan yang masih timpang. Secara agregat, performa terlihat tinggi pada akurasi, namun performa rata rata antar kelas belum merata. Ringkasan performa *baseline*:

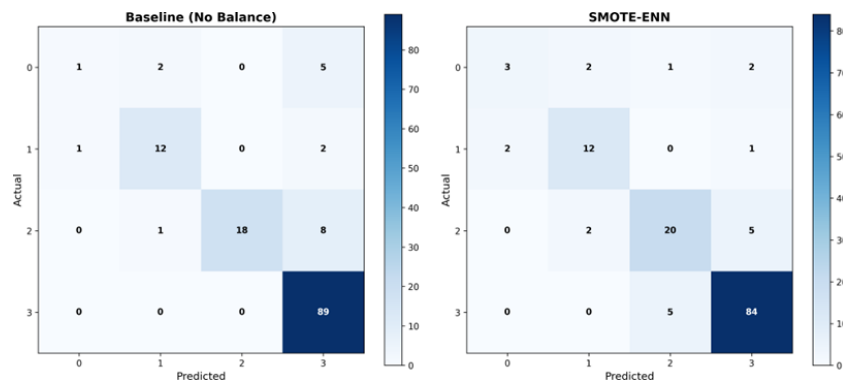
1. Akurasi 0,8633.
2. Presisi rata rata antar kelas 0,7889.
3. *Recall* rata rata antar kelas 0,6479.
4. *F1 Score* rata rata antar kelas 0,6806.
5. MAE 0,2230.

Jika dilihat per kelas, kelas 3 sangat mudah dikenali, tetapi kelas 0 sebagai kelas paling sedikit justru paling sering salah. Ringkasan per kelas pada *baseline* dapat dilihat pada Tabel 2.

Tabel 2. Performa per kelas *Recall* dan *F1-Score*

Kelas	Recall	F1-Score
0	0,12	0,20
1	0,80	0,80
2	0,67	0,80
3	1,00	0,92

Gambar 2 memperlihatkan pola kesalahan prediksi yang terjadi



Gambar 2. Perbandingan *confusion matrix baseline* dan SMOTE ENN

Dari *confusion matrix baseline*, terlihat bahwa pada kelas 0 hanya 1 data yang diprediksi benar dari total 8 data uji. Sebagian besar data kelas 0 justru bergeser ke kelas 3. Pola ini umum terjadi ketika model belajar dari data yang sangat didominasi satu kelas, sehingga keputusan model cenderung condong ke kelas mayoritas.

C. Hasil skenario dengan SMOTE ENN

Pada skenario kedua, SMOTE ENN diterapkan pada data pelatihan saja. Dampaknya terlihat jelas pada distribusi kelas di data pelatihan yang menjadi jauh lebih seimbang dan ukuran data pelatihan meningkat. Ukuran data pelatihan berubah dari 552 menjadi 1309 sampel.

Tabel 3. Distribusi kelas sebelum dan sesudah SMOTE ENN pada data pelatihan

No	Label	Sebelum	Sesudah
1	0	38	355
2	1	77	349
3	2	132	325
4	3	444	280

Visual distribusi kelas dapat dilihat pada Gambar 3.



Gambar 3. Distribusi kelas data asli dan data pelatihan sebelum dan sesudah SMOTE ENN

Setelah pelatihan ulang, metrik agregat menunjukkan perubahan yang menarik. Akurasi sedikit turun, namun metrik yang sensitif terhadap kelas minoritas justru membaik. Ringkasan performa dengan SMOTE ENN:

Tabel 4. Ringkasan Metrik Evaluasi Keseluruhan Model

Metrik	Nilai
Akurasi	0,8561
Presisi rata-rata antar kelas	0,7581
Recall rata-rata antar kelas	0,7149
F1-Score rata-rata antar kelas	0,7297
MAE	0,1871

Perubahan paling terasa terjadi pada kelas minoritas, terutama kelas 0. Perubahan per kelas yang menonjol:

Tabel 5. Perbandingan performa per kelas sebelum dan sesudah SMOTE ENN

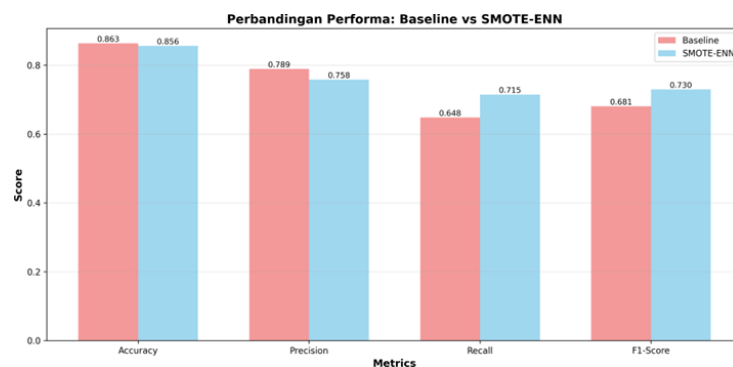
Kelas	Metrik	Sebelum	Sesudah	Perubahan	Catatan
0	Recall	0,12	0,38	+0,26	—
0	F1-Score	0,20	0,46	+0,26	—
2	Recall	0,67	0,74	+0,07	—
3	Recall	1,00	0,94	-0,06	Presisi meningkat karena model tidak lagi terlalu agresif memasukkan banyak data ke kelas 3

D. Perbandingan langsung *baseline* dan SMOTE ENN

Berikut ringkasan perbandingan metrik utama yang juga sesuai dengan visual grafik metrik.

Tabel 6. Perbandingan metrik agregat

Metrik	Baseline (No Balance)	SMOTE-ENN	Perubahan
Accuracy	0.8633	0.8561	-0.0072 (-0.83%)
Precision	0.7889	0.7581	-0.0309 (-3.91%)
Recall	0.6479	0.7149	0.0670 (+10.34%)
F1-Score	0.6806	0.7297	0.0491 (+7.21%)
MAE	0.2230	0.1871	-0.0360 (-16.13%)



Gambar 4. Perbandingan metrik *baseline* dan SMOTE ENN

E. Diskusi

Hasil eksperimen menunjukkan bahwa *baseline* menghasilkan akurasi tinggi karena model diuntungkan oleh dominasi kelas 3. Ini terlihat dari *recall* kelas 3 yang sempurna pada *baseline*. Namun, ketika fokus evaluasi diperluas ke rata-rata antar kelas, terutama kelas minoritas, *baseline* memiliki kelemahan yang jelas. Kelas 0 hampir tidak terdeteksi, yang berisiko jika sistem digunakan untuk evaluasi kualitas secara adil.

Penerapan SMOTE ENN memperbaiki masalah ini dengan dua cara. Pertama, SMOTE memperkaya representasi kelas minoritas sehingga model mendapat lebih banyak contoh untuk belajar. Kedua, ENN membersihkan sampel yang ambigu sehingga batas antar kelas menjadi lebih jelas. Kombinasi ini membuat model lebih berani memprediksi kelas minoritas, yang tercermin dari kenaikan *recall* dan F1 *Score* rata-rata antar kelas.

Akurasi dan presisi rata-rata antar kelas sedikit turun setelah SMOTE ENN. Ini adalah trade off yang wajar, karena model tidak lagi bias ke kelas mayoritas. Dalam konteks data tidak seimbang, penurunan kecil pada akurasi sering dianggap dapat diterima bila diikuti kenaikan *recall* dan F1 *Score*, karena tujuan utama adalah memperbaiki performa lintas kelas, bukan sekadar menang pada kelas mayoritas.

Penurunan MAE juga penting. Karena label bersifat bertingkat dari 0 sampai 3, MAE yang lebih kecil berarti prediksi model lebih dekat dengan label sebenarnya. Ini mengindikasikan bahwa setelah penyeimbangan, kesalahan model cenderung tidak terlalu jauh melompat antar tingkat kualitas.

F. Analisis error kualitatif

Untuk melengkapi evaluasi kuantitatif, dilakukan analisis error secara kualitatif dengan meninjau contoh proposisi pada data uji yang salah diprediksi pada *baseline* tetapi berhasil diprediksi benar setelah penerapan SMOTE ENN. Analisis ini membantu menjelaskan mengapa model *baseline* gagal dan bagaimana penyeimbangan data mengubah batas keputusan.

Tabel 7. Analisis error kualitatif

Proposisi (teks)	Label benar	Prediksi Baseline	Prediksi SMOTE-ENN	Catatan singkat
kolom milik indeks	1	3	1	teks sangat pendek
bahasa formal sebut bahasa komersial	0	3	0	kosakata jarang/ambigu
relasi milik foreign key	1	3	1	term teknis spesifik
domain rupa kumpul nilai valid	2	3	2	frasa ambigu

Secara umum, kesalahan *baseline* cenderung terjadi ketika proposisi mengandung kosakata yang jarang muncul atau term teknis yang tidak cukup terwakili pada kelas minoritas, sehingga prediksi mudah tertarik ke kelas mayoritas. Setelah SMOTE ENN, representasi kelas minoritas bertambah dan sebagian sampel ambigu dibersihkan oleh ENN, yang membantu model mengenali pola kelas minoritas dengan lebih baik.

G. Keterbatasan dataset dan arah lanjutan

Jumlah data yang digunakan (691 proposisi) relatif terbatas dan berasal dari satu sumber penilaian pakar, sehingga generalisasi hasil masih memiliki keterbatasan. Kinerja model, khususnya pada kelas minoritas, berpotensi berubah apabila dataset diperluas atau mencakup variasi topik, angkatan, maupun penilai. Sebagai pekerjaan lanjutan, disarankan melakukan evaluasi Repeated Stratified k-Fold Cross-Validation untuk mengukur stabilitas performa, menyusun learning curve untuk melihat konsistensi ketika jumlah data bertambah, serta memperluas dataset dan mempertimbangkan multi-rater labeling untuk mengurangi bias penilai tunggal.

Kesimpulan

Penelitian ini menunjukkan bahwa penerapan SMOTE ENN pada data pelatihan berhasil mengurangi bias model SVM terhadap kelas mayoritas dan meningkatkan kemampuan model dalam mengenali kelas minoritas. Hal ini terlihat dari kenaikan *recall* rata-rata antar kelas dari 0,6479 menjadi 0,7149 serta peningkatan F1 *Score* rata-rata antar kelas dari 0,6806 menjadi 0,7297. Perbaikan ini juga tercermin pada kelas 0 yang sebelumnya

sulit terdeteksi, di mana *recall* meningkat dari 0,12 menjadi 0,38 dan *F1 Score* meningkat dari 0,20 menjadi 0,46.

Meskipun akurasi sedikit menurun dari 0,8633 menjadi 0,8561 dan presisi rata-rata antar kelas juga turun, penurunan tersebut merupakan trade off yang wajar ketika model tidak lagi terlalu condong ke kelas 3. Selain itu, nilai MAE turun dari 0,2230 menjadi 0,1871, yang mengindikasikan prediksi model menjadi lebih dekat dengan label sebenarnya pada skala kualitas yang bertingkat.

Secara keseluruhan, SMOTE ENN dapat dianggap efektif untuk meningkatkan keadilan performa klasifikasi pada dataset peta konsep yang tidak seimbang, terutama ketika tujuan evaluasi tidak hanya menekankan akurasi, tetapi juga pemerataan kinerja antar kelas

Daftar Pustaka

- [1] D. D. Prasetya, A. Pinandito, Y. Hayashi, and T. Hirashima, "Analysis of quality of knowledge structure and students' perceptions in extension concept mapping," *Res. Pract. Technol. Enhanc. Learn.*, vol. 17, no. 1, p. 14, Dec. 2022, doi: 10.1186/s41039-022-00189-9.
- [2] C. Cischke and S. T. Mueller, "Concept Mapping Assessments as a Tool for Judgment of Learning," Jul. 29, 2022. doi: 10.31234/osf.io/69bjx.
- [3] F. Kamalov, S. E. Choutri, and A. F. Atiya, "Analytical formulation of synthetic minority oversampling technique (SMOTE) for imbalanced learning," *Gulf Journal of Mathematics*, vol. 19, no. 1, pp. 400–415, Jan. 2025, doi: 10.56947/gjom.v19i1.2639.
- [4] J. Wang and N. Awang, "A Novel Synthetic Minority Oversampling Technique for Multiclass Imbalance Problems," *IEEE Access*, vol. 13, pp. 6054–6066, 2025, doi: 10.1109/ACCESS.2025.3526673.
- [5] Y.-C. Wang and C.-H. Cheng, "A multiple combined method for rebalancing medical data with class imbalances," *Comput. Biol. Med.*, vol. 134, p. 104527, Jul. 2021, doi: 10.1016/j.compbiomed.2021.104527.
- [6] A. Gupta and S. Gupta, "Enhanced Classification of Imbalanced Medical Datasets using Hybrid Data-Level, Cost-Sensitive and Ensemble Methods," *International Research Journal of Multidisciplinary Technovation*, pp. 58–76, Apr. 2024, doi: 10.54392/irjmt2435.
- [7] I. D. Mienye and Y. Sun, "Performance analysis of cost-sensitive learning methods with application to imbalanced medical data," *Inform. Med. Unlocked*, vol. 25, p. 100690, 2021, doi: 10.1016/j.imu.2021.100690.
- [8] L. Sun, Z. Shang, Q. Cao, K. Chen, and J. Li, "Electrocardiogram diagnosis based on SMOTE+ ENN and random forest," in *International Conference on Intelligent Computing*, 2019, pp. 747–757.
- [9] D. Yilmaz Eroglu and M. S. Pir, "Hybrid Oversampling and Undersampling Method (HOUM) via Safe-Level SMOTE and Support Vector Machine," *Applied Sciences*, vol. 14, no. 22, p. 10438, Nov. 2024, doi: 10.3390/app142210438.
- [10] S. M. M. Hossain, K. M. A. Kamal, A. Sen, and I. H. Sarker, "TF-IDF feature-based spam filtering of mobile SMS using a machine learning approach," in *Applied Intelligence for Industry 4.0*, Chapman and Hall/CRC, 2023, pp. 162–175.
- [11] A. Ranjan, D. Fernandez-Baca, S. Tripathi, and A. Deepak, "An Ensemble Tf-Idf Based Approach to Protein Function Prediction via Sequence Segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 19, no. 5, pp. 2685–2696, Sep. 2022, doi: 10.1109/TCBB.2021.3093060.
- [12] X. Ye, Z. Zhang, and Y. Jiang, "Prediction of Breast Cancer of Women Based on Support Vector Machines," in *Proceedings of the 2020 4th International Conference on Electronic Information Technology and Computer Engineering*, New York, NY, USA: ACM, Nov. 2020, pp. 780–784. doi: 10.1145/3443467.3443853.
- [13] D. D. Prasetya, T. Widiyaningtyas, and T. Hirashima, "Interrelatedness patterns of knowledge representation in extension concept mapping," *Res. Pract. Technol. Enhanc. Learn.*, vol. 20, p. 009, May 2024, doi: 10.58459/rptel.2025.20009.
- [14] A. Novanto, D. Indra, and W. Astuti, "Analisis Pre-processing Sentimen Terhadap Komentar Layanan Indihome Pada Twitter," *LINIER: Literatur Informatika dan Komputer*, vol. 1, no. 2, pp. 145–152, Jun. 2024, doi: 10.33096/linier.v1i2.2491.
- [15] K. Yusupov, M. R. Islam, I. Muminov, M. Sahlabadi, and K. Yim, "Comparative Analysis of Machine Learning and Deep Learning Models for Email Spam Classification Using TF-IDF and Word Embedding Techniques," 2025, pp. 114–122. doi: 10.1007/978-3-031-76452-3_11.

- [16] C. A. Nurhaliza Agustina, R. Novita, Mustakim, and N. E. Rozanda, "The Implementation of TF-IDF and Word2Vec on Booster Vaccine Sentiment Analysis Using Support Vector Machine Algorithm," *Procedia Comput. Sci.*, vol. 234, pp. 156–163, 2024, doi: 10.1016/j.procs.2024.02.162.
- [17] G. Popoola, K.-K. Abdullah, G. S. Fuhnwi, and J. Agbaje, "Sentiment Analysis of Financial News Data using TF-IDF and Machine Learning Algorithms," in *2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC)*, IEEE, Feb. 2024, pp. 1–6. doi: 10.1109/ICAIC60265.2024.10433843.
- [18] A. A. G. W. S. Erlangga, I. G. A. Gunadi, and I. M. G. Sunarya, "Kombinasi Oversampling dan Undersampling dalam Menangani Class Imbalanced dan Overlapping pada Klasifikasi Data Bank Marketing," *Jurnal RESISTOR (Rekayasa Sistem Komputer)*, vol. 7, no. 1, pp. 32–42, Apr. 2024, doi: 10.31598/jurnalresistor.v7i1.1515.
- [19] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information*, vol. 14, no. 1, p. 54, Jan. 2023, doi: 10.3390/info14010054.
- [20] G. Husain *et al.*, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," *Algorithms*, vol. 18, no. 1, p. 37, Jan. 2025, doi: 10.3390/a18010037.