PENERAPAN *NATURAL LANGUAGE PROCESSING* DAN *K-MEANS CLUSTERING* UNTUK MENGANALISIS PENGARUH KOMPLEKSITAS LINGUISTIK TERHADAP TINGKAT KESALAHAN SISWA

Annisa Lestari Putri[1*], Amelia Ratu Kusumawati[2]
[1]*Program Studi Statistika, Universitas Terbuka, Bandung, Indonesia*
[2]*MA Syamsul Ulum, Kota Sukabumi, Indonesia*

**Penulis korespondensi: annisalesput@gmail.com*

ABSTRAK

Kesenjangan antara kompleksitas linguistik soal matematika dan kemampuan kognitif siswa sering kali menjadi faktor penghambat dalam proses asesmen pendidikan. Penelitian ini bertujuan untuk mengidentifikasi pengaruh kompleksitas bahasa terhadap performa siswa dan memetakan karakteristik butir soal menggunakan pendekatan *Natural Language Processing* (NLP) dan algoritma K-Means Clustering. Data penelitian diperoleh dari hasil ujian matematika 41 siswa yang kemudian diintegrasikan dengan variabel skor kompleksitas bahasa (SKB) berdasarkan tokenisasi teks dan klasifikasi tingkat kognitif taksonomi Bloom. Hasil analisis menunjukkan bahwa soal-soal dengan volume teks tinggi (SKB > 140 kata) berkorelasi signifikan dengan peningkatan error rate secara drastis yang mencapai rata-rata 66,64%. Pemodelan klasterisasi berhasil mengidentifikasi adanya fenomena titik buta sintaksis pada soal dengan level kognitif rendah (C1) yang emiliki tingkat kesalahan tidak wajar hingga 39,02% akibat beban sintaksis yang berlebihan. Temuan ini menyimpulkan bahwa kompleksitas bahasa dalam instrumen soal merupakan determinan krusial yang dapat mendistorsi kemampuan matematis siswa. Kerangka kerja evaluasi berbasis data melalui integrasi NLP sangat direkomendasikan untuk mereformasi penyusunan soal ujian agar lebih selaras dengan kapasitas kognitif dan linguistik siswa.

Kata kunci: *Natural Language Processing* (NLP), K-Means Clustering, Error Rate, Taksonomi Bloom, Titik Buta Sintaksis.

1 PENDAHULUAN

Evaluasi hasil pembelajaran merupakan proses kritis dalam pendidikan yang bertujuan untuk mengukur pencapaian tujuan pembelajaran secara komprehensif (Hendrastuty, 2024). Tingginya tingkat kesalahan siswa saat menyelesaikan soal berbasis narasi seringkali dipicu oleh beban kognitif yang berlebih, di mana beban ini sangan mempengaruhi proses kognitif dan performa siswa dalam mengerjakan soal (Noroozi & Karami, 2022). Kesulitan memecahkan masalah ini salah satunya bersumber dari pengaruh sintaksis bahasa, terutama ketika siswa dihadapkan pada struktur kalimat yang rumit dan tidak terduga pada teks soal (Dröse et al., 2021). Untuk membedah dan mengkuantifikasi elemen bahasa tersebut secara matematis, *Natural Language Processing* (NLP) menjadi unstrumen esensial. NLP merupakan cabang keilmuan yang memproses teks bahasa alami manusia secara otomatis sehingga data tidak terstruktur tersebut dapat diterjemahkan mejadi informasi yang dapat dianalisis oleh mesin (Maulud et al., 2021).

Dari sisi linguistik matematis, toleransi siswa terhadap ambiguitas bahasa terbukti memiliki hubungan positif yang signifikan terhadap kemampuan mereka dalam memecahkan masalah matematika non-rutin (Buela et al., 2020). Sementara itu, dari sisi rekayasa komputasional, metode NLP telah diimplemntasikan secara efektif untuk mengekstraksi dan memetakan soal ujian ke

dalam dimensi proses kognitif berdasarkan taksonomi Bloom (Mustafidah et al., 2022). Dalam ranah educational data mining (EDM), pemanfaatan algoritma unsupervised learning seperti K-Means clustering telah diaplikasikan secara luas untuk mengidentifikasi pola hasil belajar (Hendrastuty, 2024), serta memetakan tingkat kesulitan belajar secara objektif sebagai dasar evaluasi akademik (Saskia et al., 2026).

Meskipun kajian mengenai NLP dan algoritma K-Means telah banyak dilakukan, mayoritas penerapan NLP dalam analisis soal ujian saat ini difokuska pada pengklasifikasian level taksonomi kognitif saja (Mustafidah et al., 2022), tanpa mengintegrasikannya dengan bukti empiris berupa persentase tingkat kesalahan menjawab siswa. Di sisi lain, riset-riset klasterisasi berbasis EDM umumnya hanya membagi kelompok berdasarkan riwayat nilai akademik akhir atau interaksi pembelajaran (Bunkar & Tanwani, 2020; Saskia et al., 2026), dan cenderung mengabaikan fitur linguistik sebagai variabel penentu. Padahal, kompleksitas struktur bahasa dapat memicu munculnya titik buta sintaksis, sebuah kondisi di mana kesalahan penalaran matematis terjadi bukan karena defisit pemahaman konsep, melainkan akibat mispersepsi terhadap struktur susunan kata (Williamson et al., 2025). Hingga saat ini, belum ada kerangka kerja yang mengintegrasikan ekstraksi kompleksitas linguistik dengan klasterisasi performa siswa.

Penelitian ini bertujuan untuk menjembatani kesenjangan tersebut dengan menggunakan pendekatan evaluasi instrumen yang menggabungkan linguistik komputasional dan statistika multivariat. Fokus utama penelitian ini adalah menganalisis butir soal ujian matematika dengan memetakannya ke dalam klaster berdasarkan tingkat kompleksitas linguistik, dimensi kognitif taksonomi Bloom, dan tingkat kesalah empiris siswa. Melalui metode ini, penelitian diharapkan mampu mengisolasi anomali soal ujian yang memiliki bias bahas, sehingga dapat memberikan kerangka analitik berbasis data bagi pendidik untuk merevisi instrumen asesmen kuantitatif yang presisi dan berkeadilan.

2 METODE

2.1 Data dan Sumber Data

Data yang digunakan dalam penelitian ini mencakup dua komponen utama, yaitu data tidak tersturktur berupa teks narasi instrumen soal ujian sebanyak 10 butir soal dan data terstuktur berupa log hasil respon jawaban dari total 41 siswa yang mengikuti ujian. Sumber data penelitian ini diperoleh secara sekunder dari dokumen naskah ujian Sumatif Akhir Semester (SAS) genap mata pelajaran matematika kelas X di sebuah MA di Kota Sukabumi.

Penggunaan kombinasi data teks instrumen dan log performa siswa ini merepresentasikan kerangka kerja educational data mining untuk mengevaluasi efektivitas hasil pembelajaran secara objektif. Seluruh data teks dari narasi soal diekstraksi untuk dianalisis fitur kebahasaannya, sedangkan data respons siswa yang mencakup pilihan jawaban dari kelas sampe digunakan sebagai basis empiris dalam mengukur tingkat kesalahan.

2.2 *Error Rate*

Pengukuran hasil belajar merupakan tahapan esensial untuk menilai sejauh mana siswa mampu mencapai target kognitif yang diujikan melalui instrumen asesmen (Anderson et al., 2002). Metrik empiris berupa error rate digunakan untuk mengukur hal tersebut secara kuantititatif. Error rate merupakan metrik evaluasi standar yang digunaka untuk menghitung proporsi tingkat kesalahan dari total observasi yang ada (Eisenstein, 2018).

Data mentah berupa log jawaban siswa harus ditransformasi telebih dahulu untuk mendapatkan error rate. Proses perhitungan dilakukan melalui manipulasi data terstruktur, di mana

dilakukan agregasi frekuensi jawaban salah yang kemudian dibagi dengan total populasi sampai menghasilkan prosporsi atau persentase yang valid (Wickham & Grolemund, 2017).

Secara matematis, perhitungan kesalahan pada setiap butir soal diformulasikan sebagai berikut:

$$ER_i = \frac{f_{salah,i}}{N} \times 100\%$$

dengan:

ER_i : error rate empiris pada butir soal ke-i.
$f_{salah,i}$: frekuensi siswa yang memberikan jawaban salah pada butir soal ke-i.
N : jumlah keseluruhan siswa yang mengerjakan butir soal tersebut.

Nilai error rate yang diperoleh dari transformasi data ini akan berada pada rentang 0% sampai 100%. Skor yang mendekati 100% merepresentasikan tingkat kesulitan atau kegagalan yang ekstrem pada soal tersebut. Metrik numerik yang dihasilkan ini selanjutnya akan distandarisasi dan dijadikan sebagai salah satu variabel input utama dalam algoritma unsupervised learning untuk memetakan karakteristik soal ke dalam klaster-klaster spesifik (Kassambara, 2017).

2.3 Skor Kompleksitas Bahasa

Komplesitas bahasa dinilai melalui panjang pendeknya unit kalimat serta volume korpos kata yang menyusun suatu dokumen (Eisenstein, 2018). Proses penambangan teks inidiimplementasikan menggunakan pendekatan tidy text di Rstudio melalui tahapan tokenization, yaitu pemecahan baris narasi soal menjadi unit kata tunggal atau token tanpa mengubah konteks utamanya (Silge & Robinson, 2017).

Skor kompleksitas bahasa (SKB) didefinisikan berdasarkan akumulasi jumlah token kata yang membangun satu kesatuan unit soal. Formulasi fitur kuantitatif kompleksitas ini ditulis secara matematis sebagai berikut:

$$SKB_i = \sum_{j=1}^{M_i} w_{j,i}$$

dengan:

SKB_i : skor kompleksitas bahasa pada butir soal ke-i.
$w_{j,i}$: token kata ke-j yang menyusun komponen teks pada butir soal ke-i.
M_i : jumlah total keseluruhan token kata yang terdeteksi pada butir soal ke-i.

Metrik numerik SKB merepresentasikan dimensi volume teks. Semakin tinggi nilai SKB, semakin panjang narasi yang harus diproses siswa sebelum dapat melakukan penalaran matematis. Nilai ini selanjutnya digunakan sebagai salah satu variabel input utama dalam algoritma klasterisasi (Silge & Robinson, 2017).

2.4 Klasifikasi Berdasarkan Taksonomi Bloom

Setiap butir soal ujian diklasifikasikan ke dalam dimensi proses kognitif untuk mengukut kedalaman kemampuan berpikir yang dituntut dari siswa. Kerangka klasifikasi ini mengacu pada taksonomi Bloom yang telah direvisi.

Tabel 1. Hierarki Taksonomi Bloom Versi Revisi

Tingkatan	Jenis kemampuan
C1	Mengingat
C2	Memahami
C3	Menerapkan
C4	Menganalisis
C5	Mengevaluasi
C6	Membuat

Sumber: (Anderson *et al*., 2002)

Tingkatan ini secara umum dikelompokkan menjadi dua kategori utama, yaitu lower order thinking skills (LOTS) untuk ranah C1-C3 dan higher order thinking skills (HOTS) untuk ranah (C4-C6).

Proses klasifikasi teks soal dilakukan secata komputasional melalui pendekatan text mining. Algoritma NLP memindai teks pada instrumen ujian untuk mengidentifikasi kata kunci atau kata kerja operasional (KKO) utama yang bertindak sebagai indikator instruksi kognitif (Anderson et al., 2002). Hasil pelabelan kategori kognitif C1 hingga C6 ini kemudian ditransformasikan menjadi skala ordinal numerik (angka 1-6) agar dapat diolah dan diselaraskan secara statistik sebagai salah satu variabel input utama dalam analisis kuantitatif lanjutan.

2.5 Langkah Analisis

Langkah-langkah penelitian yang dilakukan dalam penelitian ini yaitu:

1. Mengkalkulasi persentase error rate siswa dan membersihkan teks dokumen soal dari karakter yang tidak bermakna (Wickham & Grolemund, 2017).
2. Memproses tokenisasi teks soal menggunakan algoritma NLP untuk mengitung skor kompleksitas bahasa (Eisenstein, 2018).
3. Memetakan instrumen soal ke dalam skala ordinal (1-6) berdasarkan panduan taksonomi Bloom versi revisi (Anderson et al., 2002).
4. Menormalisasi rentang nilai error rate, skor bahasa, dan level kognitif agar jarak metriknya seimbang sebelum diklasterisasi (Kassambara, 2017).
5. Menggunakan Elbow Method untuk menentukan jumlah partisi kelompok (K) yang paling ideal bagi himpunan data (Kassambara, 2017).
6. Mengidentifikasi klaster soal yang terindikasi memiliki bias bahasa ekstrem sebagai dasar rekomendasi revisi instrumen (Wickham & Grolemund, 2017).

3 HASIL DAN PEMBAHASAN

3.1 Perhitungan Error Rate

Berikut adalah hasil perhitungan error rate bagi setiap butir soal.

Tabel 2. Error Rate untuk Setiap Butir Soal

Soal ke-i	Jumlah siswa menjawab salah (f_{salah})	Error rate (%)
1	16	39,02%
2	20	48,78%
3	23	56,10%
4	32	78,05%

Soal ke-i	Jumlah siswa menjawab salah (f_{salah})	Error rate (%)
5	34	82,93%
6	8	19,51%
7	11	26,83%
8	8	19,51%
9	10	24,39%
10	7	17,07%

Data pada Tabel 2 memperlihatkan adanya variasi tingkat kesulitan empiris yang sangat kontras di antara butir-butir soal yang diujikan. Analisis deskriptif menunjukkan bahwa kelompok soal nomor 1 sampai 5 memiliki kecenderungan error rate yang jauh lebih tinggi dibandingkan kelompok soal nomor 6 sampai 10. Tingkat kegagalan tertinggi ditemukan pada soal nomor 5 dengan error rate ekstrem sebesar 82,93%, disusul oleh soal nomor 4 sebesar 78,05%. Sedangkan tingkat kegagalan terendah tercatat pada soal nomor 10 dengan nilai error rate hanay 17,07%, diikuti oleh soal 6 dan 8 yang masing-masing memiliki error rate sebesar 19,51%.

Ketimpangan performa siswa antar nomor soal ini mengindikasikan adanya perbedaan karakteristik parameter intrinsik pada butir soal, yang selanjutnya akan dianalisis lebih lanjut melalui perhitungan skor kompleksitas bahasa.

Tabel 3. Hasil Ekstraksi Skor Kompleksitas Bahasa (SKB) Berbasis Tokenisasi NLP

Soal ke-i	Kalimat pertanyaan	Total kata (SKB)	Kategori volume teks
1	Definisi ukuran pemusatan	163	Tinggi
2	Analisis sifat distribusi	193	Sangat tinggi
3	Perhitungan rata-rata	149	Tinggi
4	Penentuan nilai median	173	Sangat tinggi
5	Proyeksi rata-rata baru	170	Sangat tinggi
6	Kesesuaian diagram batang	84	Rendah
7	Kesesuaian diagram garis	83	Rendah
8	Kesesuaian diagram lingkaran	79	Rendah
9	Analisis sektor diagram	81	Rendah
10	Analisis urutan diagram	85	Rendah

Tabel 3 menunjukkan adanya dikotomi volume teks yang sangat jelas pada instrumen ujian yang digunakan. Kelompok soal nomor 1 sampai 5 memiliki SKB yang sangat padat, berkisar antara 149-193 kata per soal. Sebaliknyam kelompok soal nomor 6 sampai 10 disajikan dengan kalimat yang lebih ringkas, dengan skor SKB ada di antara rentang 79-85 kata.

Perbedaan ekstrem dalam volume token kata ini mengonfirmasi bahwa tidak semua butir soal matematika dalam satu instrumen ujian memberikan beban bacaan yang seragam. Soal-soal dengan nilai SKB di atas 140 kata menuntut kapasitas memori yang lebih besar dari siswa untuk menyaring informasi numerik dari tumpukan test naratif (Eisenstein, 2018). Tingginya beban linguistik pada paruh pertama soal ujian in inantinya akan diselaraskan dengan parameter tingkat kognitif dan error rate untuk membuktikan ada atau tidaknya distorsi kebahasaan dalma instrumen evaluasi tersebut.

3.2 Perhitungan Skor Kompleksitas Bahasa

Tahap perhitungan skor kompleksitas bahasa melibatkan proses kuantifikasi struktur narasi kulaitatif dari instrumen ujian menjadi metrik numerik yang terukur. Setiap kalimat dalam narasi dan pertanyaan dipecah menjadi unit kata tunggal menggunakan teknik NLP melalui kerangka kerja tidy text. Total akumulasi token kata pada setiap nomor soal kemudian diekstrak sebagai representasi skor kompleksitas bahasa (SKB). Metrik ini berfungsi sebagai indikator objektif terhadap besarnya beban sintaksis atau volume bacaan yang harus diproses oleh siswa sebelum mereka dapat melakukan penalaran matematis (Silge & Robbinson, 2017).

3.3 Pengklasifikasian Berdasarkan Taksonomi Bloom

Dimensi proses kognitif merupakan variabel fundamental yang mempengaruhi performa siswa dalam menjawab instrumen asesmen. Berdasarkan Anderson et al. (2002), instruksi pada setiap butir soal dianalisis berdasarkan kata kerja operasional (KKO) yang digunakan. Hal ini bertujuan untuk memetakan beban pemikiran matematis aktual yang dituntut oleh soal terlepas dari panjang atau pendeknya narasi teks yang disajikan.

Tabel 4. Hasil Klasifikasi Tingkat Kognitif Soal Berdasarkan Taksonomi Bloom Versi Revisi

Soal ke-i	Identifikasi KKO	Dimensi kognitif
1	"…secara definisi disebut sebagai…"	Mengingat (C1)
2	"…pernyataan yang tepat mengenai sifat…"	Menganalisis (C4)
3	"…nilai rata-rata hitung…adalah…"	Menerapkan (C3)
4	"…nilai median…adalah…"	Menerapkan (C3)
5	"…nilai berat badan siswa yang baru…haruslah sebesar…"	Menganalisis (C4)
6	"…diagram yang memiliki…tersebut adalah…"	Menerapkan (C3)
7	"…bentuk diagram yang paling tepat digunakan…"	Menerapkan (C3)
8	"Bentuk diagram lingkaran yang sesuai…"	Menerapkan (C3)
9	"Diagram lingkaran yang memiliki sektor…"	Menerapkan (C3)
10	"Bentuk diagram barang yang sesuai dengan urutan…"	Menerapkan (C3)

Tabel 4 menunjukkan distribusi instrumen soal didominasi oleh level C3 (menerapkan) sebanyak 7 soal. Pada level ini, siswa dituntut untuk menggunakan prosedur atau rumus matematika pada situasi yang diberikan (Anderson et al., 2002). Sementara itu soal 2 dan 5 masuk ke dalam kategori higher order thinking skills (HOTS) pada level C4, di mana siswa harus memecah materi ke dalam bagian-bagian komponennya dan menentukan bagaimana bagian-bagian tersebut saling berhubungan.

Terdapat temuan menarik pada soal nomor 1 yang hanya menuntut kemampuan C1. Secara teoretis, soal pada level C1 (lower order thinking skills/LOTS) menuntut beban kognitif yang

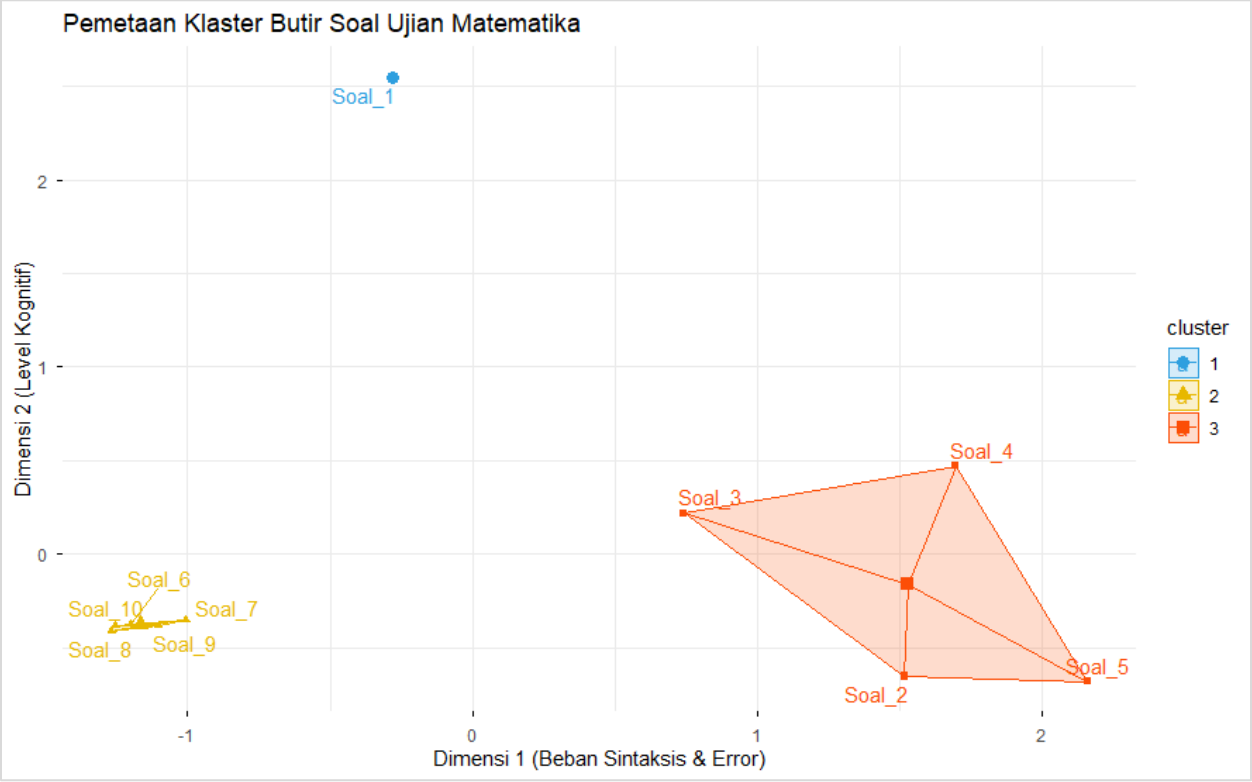
paling rendah karena hanya meminta siswa memanggil kembali memori tentang sebuah definisi faktual (Anderson et al., 2002). Dalam kondisi ideal, error rate pada tipe soal ini seharusnya sangat rendah. Namun sebagaimana hasil ekstraksi pada Tabel 3, soal 1 memiliki SKB yang tergolong tinggi yaitu 163 kata.

3.4 Pembahasan

Sebelum dimasukkan ke dalam algoritma unsupervised learning, error rate, SKB, dan skala kognitif Bloom distandarisasi untuk mencegah variabel dengan rentang nilai besar mendominasi variabel berskala kecil dalam perhitungan jarak Euclidean (Kassambara, 2017). Berdasarkan evaluasi varians intrakelompok menggunakan metode Elbow, jumlah klaster optimal yang merepresentasikan karakteristik data ditetapkan sebanyak 3 kelompok (K = 3).

Tabel 5. Profil Rata-Rata (Sentroid) Klaster Butir Soal Ujian

Klaster	Anggota soal (soal ke-i)	Rata-rata error rate (%)
Klaster 1	6, 7, 8, 9, 10	21,46
Klaster 2	2, 3, 4, 5	66,46
Klaster 3	1	39,02



Gambar 1. Visualisasi Sebaran Klaster Butir Soal Matematika

Klaster 1 berisi setengah dari total soal ujian. Kelompok ini menunjukkan karakteristik instrumen evaluasi matematika yang sangat sehat. Dengan rata-rata panjang teks yang tingkas (82,4 kata) dan tuntutan kognitif untuk menerapkan rumus (C3), mayoritas siswa mampu menangkap instruksi dengan baik sehingga error rate-nya sangat rendah (21,46%). Hal ini membuktikan bahwa ketika beban sintaksis bahasa ditekan seminimal mungkin, performa empiris siswa dapat merepresentasikan kemampuan matematis mereka yang sesungguhnya secara lebih murni.

Klaster 2 memuat soal nomor 2, 3, 4, dan 5. Siswa menghadapi dua tantangan berat sekaligus pada kelompok ini dengan level kognitif tingkat menengah-tinggi (C3-C4) serta beban membaca narasi panjang yang ekstrem (rata-rata 171,25 kata). Kombinasi antara komleksitas linguistik dan kompleksitas matematis ini menghasilkan beban kognitif ganda, yang berujung pada tingginya angka error rate secara drastis hingga mencapai rata-rata 66,46%.

Temuan paling krusial terdapat pada klaster 3 yang hanya beranggotakan soal nomor 1. Secara taksonomi kognitif, soal ini merupakan soal yang paling mudah di seluruh instrumen ujian karena hanya menuntut kemampuan mengingat definisi dasar (C1). Dalam teori pengujian klasik, error rate soal dengan tipe C1 seharusnya berada di bawah 10%. Namun pada kenyataannya, tingkat kegagalan empiris pada soal ini melonjak hingga mendekati 40%.

Anomali ini dapat dijelaskan secara komputasional melalui nilai SKB-nya yang menembus angka 163 kata. Soal 1 terindikasi mengalami fenomena titik buta sintaksis, di mana struktur bahasa yang terlalu berbelit-belit mengaburkan substansi pertanyaan utama yang sebenarnya sangat sederhana. Kegagalan siswa dalam menjawab soal ini besar kemungkinan bukan disebabkan oleh ketidakmampuan mereka dalam mengingat konsep matematika, melainkan karena beban kognitif yang terlalu tinggi saat memproses tata bahasa instruksi yang tidak selaras dengan tingkat kesulitan substansinya.

4 KESIMPULAN

Penelitian ini berhasil mengungkapkan bahwa kompleksitas bahasa dalam instrumen soal matematika memainkan peranan krusial yang sering kali terabaikan dalam evaluasi pendidikan. Analisis klasterisasi membuktikan adanya titik buta sintaksis, fenomena anomali pada soal dengan level kognitif rendah yang justru menunjukkan tingkat kesalahan tinggi akibat beban sintaksis yang berlebihan. Temuan ini relevan sebagai kerangka kerja evaluasi berbasis data untuk mereformasi penyusunan soal ujian agar beban kognitif siswa lebih difokuskan pada penalaran matematis murni tanpa terdistorsi oleh hambatan linguistik yang tidak perlu.

UCAPAN TERIMAKASIH

Puji syukur dipanjatkan ke hadirat Tuhan Yang Maha Esa, karena atas segala rahmat dan karunia-Nya penelitian ini dapat diselesaikan dengan baik. Rasa terima kasih yang mendalam dan istimewa dipersembahkan kepada Ayah dan Almarhumah Ibu, sosok yang senantiasa menjadi sumber kekuatan, doa, dan inspirasi terbesar, serta yang selalu menanamkan semangat untuk terus belajar tanpa henti. Apresiasi setinggi-tingginya juga disampaikan kepada seluruh pihak yang telah memberikan masukan dan dukungan secara moril maupun materiil selama proses penyusunan penelitian ini. Semoga tulisan ini dapat memberikan manfaat dan kontribusi positif bagi perkembangan ilmu pengetahuan.

DAFTAR PUSTAKA

Anderson, L. W., Krathwohl, D. R., & Airasian, P. W. (2002). *A Taxonomy for Learning, Teching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives (Abridge Edition)*. Longman/Pearson.

Buela, M., Joaquin, M. N., Tandang, N., & Bulasag, A. (2020). Association of Ambiguity Tolerance and Problem-solving Ability of Students in Mathematics. *International Journal of Sciences, 51*(1).

Bunkar, K., & Tanwani, S. (2020). Educational Data Mining for Student Learning Pattern Analysis using Clustering Algorithms. *International Journal of Engineering and Advanced Technology*, *9*(6), 481–488. https://doi.org/10.35940/ijeat.F1528.089620

Dröse, J., Prediger, S., Neugebauer, P., Delucchi Danhier, R., & Mertins, B. (2021). Investigating Students' Processes of Noticing and Interpreting Syntactic Language Features in Word Problem Solving through Eye-Tracking. *International Electronic Journal of Mathematics Education*, *16*(1), em0625. https://doi.org/10.29333/iejme/9674

Eisenstein, J. (2018). *Introduction to Natural Language Processing*. (Original work published MIT Press)

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Kassambara, A. (2017). *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning* (1st ed.). STHDA.

Maulud, D. H., Ameen, S. Y., Omar, N., Kak, S. F., Rashid, Z. N., Yasin, H. M., Ibrahim, I. M., Salih, A. A., Salim, N. O. M., & Ahmed, D. M. (2021). Review on *Natural Language Processing* Based on Different Techniques. *Asian Journal of Research in Computer Science*, 1–17. https://doi.org/10.9734/ajrcos/2021/v10i130231

Mustafidah, H., Suwarsito, S., & Pinandita, T. (2022). *Natural Language Processing* for Mapping Exam Questions to the Cognitive Process Dimension. *International Journal of Emerging Technologies in Learning (iJET)*, *17*(13), 4–16. https://doi.org/10.3991/ijet.v17i13.29095

Noroozi, S., & Karami, H. (2022). A scrutiny of the relationship between cognitive load and difficulty estimates of language test items. *Language Testing in Asia*, *12*(1), 13. https://doi.org/10.1186/s40468-022-00163-8

Saskia, S. A., Anisa, Y. N., & Mahendra, R. Y. (2026). PEMETAAN KESULITAN BELAJAR MAHASISWA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING. *Didaktik: Jurnal Ilmiah PGSD FKIP Universitas Mandiri, 12*(01), 20–30. https://doi.org/https://doi.org/10.36989/didaktik.v12i01.11435

Silge, J., & Robinson, D. (2017). *Text Mining with R: A Tidy Approach*. O'Reilly Media.

Wickham, H., & Grolemund, G. (2017). *R for Data Science: Import, Tidy, Transfor, Visualize, and Model Data* (1st ed.). O'Reilly Media.

Williamson, D., Ji, Y., & Dwyer, M. (2025). *Syntactic Blind Spots: How Misalignment Leads to LLMs Mathematical Errors* (arXiv:2510.01831). arXiv. https://doi.org/10.48550/arXiv.2510.01831