

PREDICTING SCHOOL NUMERACY PERFORMANCE THROUGH EDUCATIONAL DATA MINING: AN INTEGRATED K-NEAREST NEIGHBOR AND MULTINOMIAL LOGISTIC REGRESSION APPROACH

Sitti Busyrah Muchsin¹, Dadang Juandi², Suhendra³, Hasan Hamid⁴

^{1,2,3}Universitas Pendidikan Indonesia, Bandung, Indonesia

⁴Universitas Khairun, Ternate, Indonesia

e-mail: dadang.juandi@upi.edu

Abstract

International large-scale assessments and Indonesia's National Assessment continue to reveal substantial variation in school numeracy performance, creating a need for analytical approaches that can identify at-risk schools while retaining educational interpretability. Prior educational data mining studies have largely focused on student-level grades or pass/fail outcomes, whereas numeracy research has primarily emphasized descriptive or explanatory associations. This study addresses that gap through a secondary-data predictive-explanatory design integrating K-Nearest Neighbor (KNN), Synthetic Minority Over-sampling Technique (SMOTE), and Multinomial Logistic Regression. School-level data were obtained from the 2023 Indonesian Education Report for 181 elementary schools in Makassar, South Sulawesi. The outcome was the observed numeracy proficiency tier, while reading literacy and three numeracy-process indicators—Knowing/Formulating, Applying/Employing, and Reasoning/Interpreting—served as predictors. Using an 80:20 train-test split, the KNN model ($k = 5$) achieved 90% overall test accuracy; however, the macro F1-score was 0.61 because the single Intervention Required case in the test set was not correctly classified. Multinomial Logistic Regression significantly improved fit over the intercept-only model ($\chi^2 = 164.602$, $df = 8$, $p < .001$), with Nagelkerke pseudo- $R^2 = .782$. All four predictors were significant, and Employing produced the largest likelihood-ratio contribution ($\chi^2 = 12.370$, $p = .002$). The findings support a complementary use of prediction and explanation for early school-level screening and targeted pedagogical reflection, while remaining predictive and associational rather than causal.

Keywords: educational data mining, k-nearest neighbor, multinomial logistic regression, numeracy performance, national assessment

Abstrak

Asesmen internasional dan Asesmen Nasional Indonesia terus menunjukkan adanya variasi yang substansial pada capaian numerasi antarsekolah, sehingga diperlukan pendekatan analitik yang mampu mengidentifikasi sekolah berisiko sekaligus tetap dapat ditafsirkan secara edukasional. Penelitian Educational Data Mining sebelumnya lebih banyak memprediksi nilai atau kelulusan pada tingkat siswa, sedangkan penelitian numerasi umumnya menekankan analisis deskriptif atau hubungan eksplanatori. Penelitian ini mengisi kesenjangan tersebut melalui desain prediktif-eksplanatori berbasis data sekunder yang mengintegrasikan K-Nearest Neighbor (KNN), Synthetic Minority Over-sampling Technique (SMOTE), dan Regresi Logistik Multinomial. Data tingkat sekolah diperoleh dari Rapor Pendidikan 2023 yang mencakup 181 sekolah dasar di Kota Makassar, Sulawesi Selatan. Variabel terikat adalah kategori capaian numerasi sekolah, sedangkan literasi membaca dan tiga indikator proses numerasi—Knowing/Formulating, Applying/Employing, dan Reasoning/Interpreting—digunakan sebagai prediktor. Dengan pembagian data latih-uji 80:20, model KNN ($k = 5$) mencapai akurasi uji keseluruhan 90%; namun macro F1-score sebesar 0,61 karena satu-satunya kasus kategori Perlu Intervensi pada data uji tidak berhasil diklasifikasikan. Regresi Logistik Multinomial meningkatkan kecocokan model secara signifikan dibanding model intercept-only ($\chi^2 = 164,602$; $df = 8$; $p < 0,001$), dengan Nagelkerke pseudo- $R^2 = 0,782$. Seluruh prediktor signifikan dan Employing memberikan kontribusi likelihood-ratio terbesar ($\chi^2 = 12,370$; $p = 0,002$). Temuan ini mendukung penggunaan prediksi dan eksplanasi secara komplementer untuk pemetaan dini sekolah dan refleksi pedagogis yang lebih terarah, dengan tetap menempatkan hasil sebagai bukti prediktif dan asosiatif, bukan kausal.

Kata kunci: educational data mining, k-nearest neighbor, regresi logistik multinomial, capaian numerasi, asesmen nasional

INTRODUCTION

Numeracy is not reducible to routine calculation; it concerns the capacity to mobilize mathematical knowledge, representations, tools, and reasoning to interpret and act in situations encountered in personal, social, and civic life (Johnston & Yasukawa, 2001; Jain & Rogers, 2019; Geiger et al., 2015). This broader conception is particularly important in contemporary mathematics education because large-scale assessments increasingly emphasize contextualized mathematical reasoning rather than decontextualized procedural fluency. PISA 2022 again positioned mathematics as a central domain and documented substantial variation in mathematical performance across education systems (OECD, 2023). In Indonesia, the National Assessment operationalizes a similar policy orientation through the Minimum Competency Assessment (Asesmen Kompetensi Minimum/AKM), which measures reading literacy and numeracy as foundational competencies intended to inform improvement at the level of the education system and school (Pusmenjar, 2020a; Rohim, 2021).

From an epistemic perspective, school numeracy performance is multidimensional. Students must first make sense of contextual information, recognize mathematically relevant quantities and relations, mobilize concepts and procedures, and then evaluate whether the resulting mathematical solution remains meaningful in the original situation. Reading literacy is therefore not merely an external correlate of numeracy; it can function as an access condition for context-rich mathematics tasks. Empirical work has shown that reading performance explains a meaningful part of variation in mathematics performance, especially when tasks contain substantial linguistic and representational demands (Ding & Homer, 2020; Stoffelsma & Spooren, 2019). Likewise, the mathematical-literacy literature emphasizes interconnected processes of formulating situations mathematically, employing mathematical concepts and procedures, and interpreting mathematical outcomes (Stacey, 2015). In the Indonesian AKM framework, the reported cognitive levels of knowing, applying, and reasoning are conceptually related to these international processes, although they should not be treated as psychometrically identical constructs. This distinction is important because the present study uses the dual labels Knowing/Formulating, Applying/Employing, and Reasoning/Interpreting as interpretive bridges rather than claims of construct equivalence.

At the same time, Educational Data Mining (EDM) has expanded the possibilities for using assessment data prospectively. Reviews of student-performance prediction show that machine-learning studies commonly focus on individual grades, pass/fail outcomes, or course performance, with classification accuracy as the primary endpoint (Zhang et al., 2021). Recent work in elementary mathematics has similarly demonstrated the feasibility of EDM for predicting academic achievement and has highlighted the methodological challenge created by imbalanced outcome classes (Tjahyadi & Tude, 2025). Such predictive models are useful for early warning, but prediction alone does not reveal which educational dimensions are most strongly associated with membership in a particular proficiency category. Conversely, conventional explanatory models can quantify associations but do not automatically provide a flexible non-parametric screening mechanism. Class imbalance further complicates school-level prediction because rare low-performance categories can be overshadowed by majority classes; SMOTE was developed precisely to improve learning from such imbalanced distributions (Chawla et al., 2002).

This creates a clear methodological and substantive gap in mathematics education research. Existing numeracy studies provide important explanations of literacy, reasoning, and contextual problem solving, while EDM studies provide increasingly sophisticated prediction of academic outcomes; however, these two analytic purposes are rarely integrated using official school-level numeracy indicators. The novelty of this study therefore lies not simply in applying two statistical techniques, but in assigning them complementary epistemic functions. KNN is used as a non-parametric classifier to examine whether schools can be screened into numeracy proficiency categories from educationally meaningful indicators, while Multinomial Logistic Regression is used to determine how strongly those indicators distinguish the observed categories without imposing a proportional-odds assumption. By combining prediction with interpretable explanation, the study reframes Education Report data from a retrospective reporting resource into a potential basis for early diagnostic reflection. Accordingly, the study asks: (1) How accurately can KNN classify elementary schools into observed numeracy proficiency categories? (2) Which school-level indicators significantly distinguish those categories? and (3) What pedagogical and methodological insights emerge when predictive and explanatory evidence are interpreted together?

METHODS

Research Design, Unit of Analysis, and Data Source

This study employed a quantitative secondary-data predictive-explanatory design rather than an experimental or instructional-intervention design. The unit of analysis was the school, not individual students. The analytic dataset consisted of 181 public and private elementary schools in Makassar, South Sulawesi, represented in the 2023 Indonesian Education Report (Rapor Pendidikan). In the observed outcome distribution, 95 schools were classified as Basic, 85 as Proficient, and 1 as Intervention Required. No schools or students were assigned to treatment and control groups, and no instructional treatment was implemented by the researchers. The term 'implementation' in this study therefore refers to the implementation of the analytic pipeline, not the implementation of a learning model. Because the data were aggregated at school level and contained no individual identifiers in the analytic file, the study did not analyze identifiable student-level records.

Measures and Measurement Quality

The study did not develop or administer a new research instrument. Instead, it analyzed school-level indicators reported in Education Report and derived from Indonesia's National Assessment. AKM is a standardized national assessment designed to measure reading literacy and numeracy through context-rich tasks. Official development procedures include framework specification, item writing, expert review, validation, field testing, scoring and item analysis, interpretation of item characteristics, and item selection before operational use (Pusmenjar, 2020a, 2020b). Independent development research on AKM-oriented numeracy instruments also supports the relevance of understanding, application, and reasoning as measurable dimensions of numeracy performance (Purnomo et al., 2022). In the present study, the predictors were treated as administrative school-level measures; item-level responses were not available. Consequently, internal-consistency reliability could not be re-estimated from the secondary dataset. This is reported explicitly rather than attributing unobserved reliability coefficients to the national instrument.

Table 1. Operationalization of the Study Variables

Variable	Role	Operational meaning in this study	Data form
Numeracy Capability Tier	Outcome	Observed school-level numeracy category reported in Education Report: Intervention Required, Basic, or Proficient.	Categorical (3 observed levels)

Variable	Role	Operational meaning in this study	Data form
Reading Literacy	Predictor	School-level indicator of capacity to understand, use, evaluate, and reflect on written information relevant to learning.	Continuous
Knowing / Formulating	Predictor	AKM knowing indicator interpreted as conceptually related to recognizing relevant mathematical information and structuring contextual situations mathematically.	Continuous
Applying / Employing	Predictor	AKM applying indicator interpreted as conceptually related to mobilizing mathematical concepts, procedures, facts, and reasoning to solve contextual problems.	Continuous
Reasoning / Interpreting	Predictor	AKM reasoning indicator interpreted as conceptually related to evaluating, extending, and interpreting mathematical conclusions in context.	Continuous

Analytical Procedure and Validity Safeguards

The analysis was conducted in two complementary stages. First, a K-Nearest Neighbor (KNN) classifier was used because it is a non-parametric, distance-based method capable of representing local and potentially non-linear structure without assuming a specific functional relationship between predictors and outcome classes. Because KNN is sensitive to feature scale, all continuous predictors were standardized before distance calculation. The dataset was partitioned into 80% training data and 20% testing data using a fixed `random_state` to support reproducibility. To reduce information leakage, SMOTE resampling was confined to the training partition; the untouched testing partition was reserved for final performance evaluation. The final model used $k = 5$ and Euclidean distance. Performance was interpreted using overall accuracy together with class-specific precision, recall, F1-score, macro-averaged F1, and the confusion matrix. This combination is important because overall accuracy can be misleading when one outcome class is extremely rare.

Second, Multinomial Logistic Regression was estimated using maximum likelihood in SPSS 26 to identify which indicators significantly distinguished the three observed numeracy categories. Although the categories are substantively ordered, multinomial rather than ordinal logistic regression was selected because the study sought category-specific contrasts without imposing the proportional-odds assumption. Model adequacy was evaluated through Pearson and Deviance goodness-of-fit statistics, the change in $-2 \text{ Log Likelihood}$ between the intercept-only and final models, and pseudo- R^2 indices. Predictor contribution was assessed with Likelihood Ratio Tests at $\alpha = .05$. Nagelkerke pseudo- R^2 was reported as an indicator of

model-level explanatory magnitude rather than as a causal intervention effect size. Conventional intervention effect sizes such as Cohen's d are not applicable because no instructional treatment or pretest-posttest comparison was conducted. Analytical validity was strengthened by separating model training from test evaluation, preserving an unseen test set for KNN, reporting minority-class performance, and interpreting all regression findings as school-level associations rather than individual-level causal effects.

RESULTS AND DISCUSSION

K-Nearest Neighbor (KNN) Predictive Performance

The 20% testing partition contained 40 schools: 16 Proficient, 23 Basic, and 1 Intervention Required. The KNN classifier correctly classified 36 of the 40 schools, yielding an overall test accuracy of 0.90. As shown in Figure 2, performance was strong for the two majority categories. The Proficient category achieved precision = 0.84, recall = 1.00, and F1 = 0.91, while the Basic category achieved precision = 0.95, recall = 0.87, and F1 = 0.91. In contrast, the single Intervention Required case was not correctly identified, producing precision, recall, and F1 values of 0.00 for that class. Consequently, the macro-averaged precision, recall, and F1 were 0.60, 0.62, and 0.61, respectively, despite the high overall accuracy. This distinction is methodologically important: the model performs well for the dominant Basic and Proficient categories in this dataset, but the evidence is insufficient to claim reliable screening of the rarest and most educationally vulnerable category.

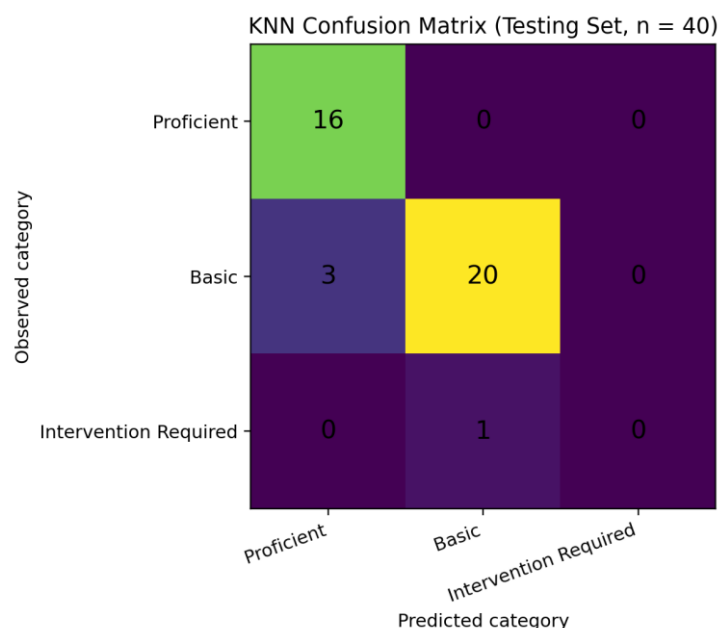


Figure 1. KNN Confusion Matrix and Classification Report on the Testing Partition

Proficient category consistently maintain significantly higher continuous coordinates across Literacy, Formulate (L1), Employ (L2), and Interpret (L3) domains compared to institutions situated in the *Basic* or *Intervention Required* categories. The visible data clustering patterns confirm that the chosen features provide solid predictive indicators, allowing the distance-based classification engine to achieve excellent performance.

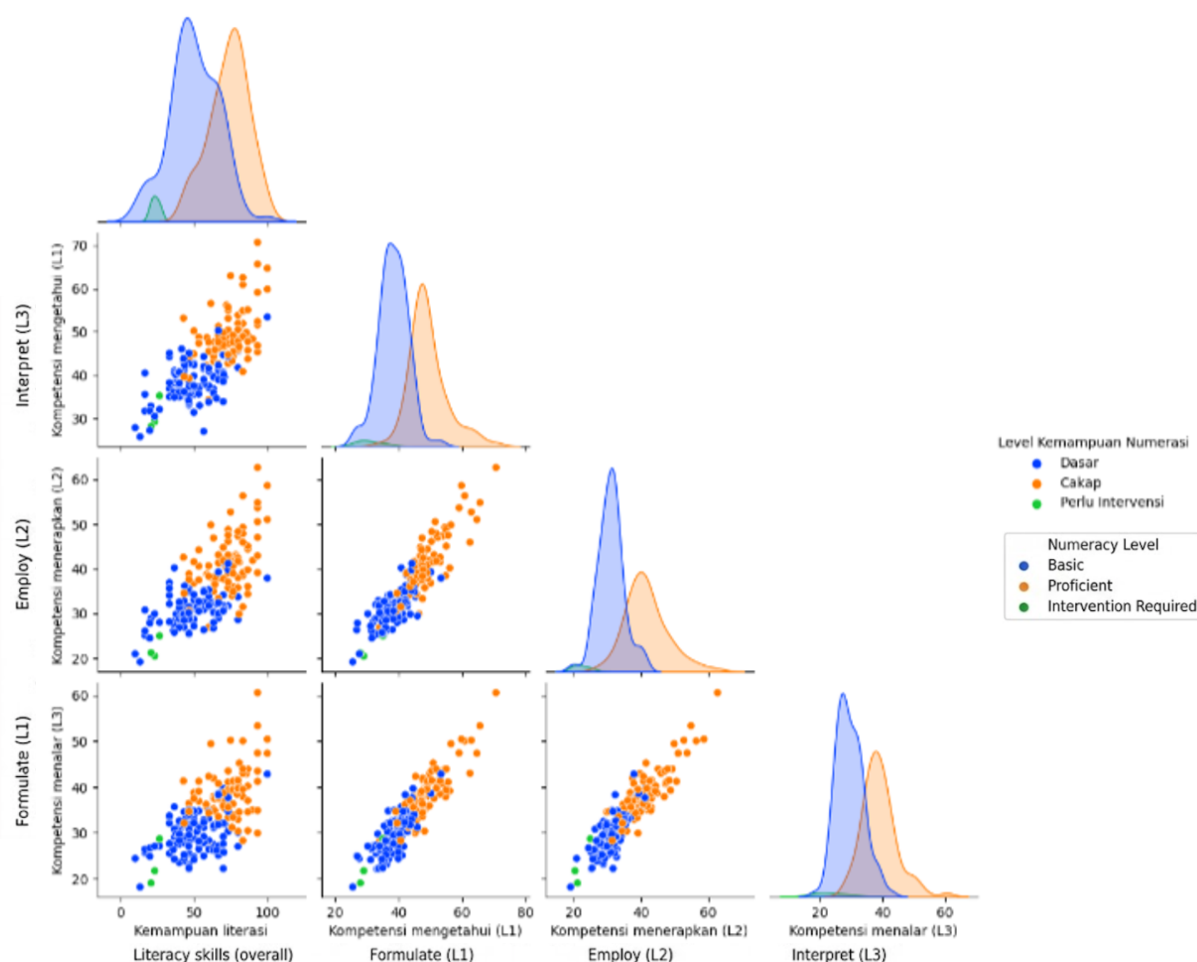


Figure 3. Pair Plot of Reading Literacy and Numeracy-Process Indicators by Observed School Numeracy Category

When the final KNN configuration was applied to the illustrative unseen school profile used in the original analysis, the profile was classified as Basic. This example demonstrates how the model can operationalize an early-warning classification for a new profile, but it should not be interpreted as external validation. External validation requires an independent dataset from a different assessment period or jurisdiction. The present result is therefore best understood as a proof of analytic usability within the available data structure.

Multinomial Logistic Regression: Model Adequacy and Explanatory Magnitude

The explanatory stage examined whether the four indicators significantly distinguished the observed numeracy categories. Goodness-of-fit statistics were non-significant for Pearson ($\chi^2 = 283.695$, $df = 346$, $p = .994$) and Deviance ($\chi^2 = 96.767$, $df = 346$, $p = 1.000$), indicating no evidence of gross lack of fit in the fitted multinomial model. The Model Fitting Information showed a substantial reduction in -2 Log Likelihood from 261.369 for the intercept-only model to 96.767 for the final model. The resulting likelihood-ratio test was significant ($\chi^2 = 164.602$, $df = 8$, $p < .001$), indicating that the predictor set improved the model relative to an intercept-only specification.

Table 2. Model Fitting Information

Model	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood	Chi-Square	Df	Sig.
Intercept Only	261.369			
Final	96.767	164.602	8	.000

To address the practical magnitude of the explanatory model, pseudo- R^2 indices were also examined. Cox and Snell R^2 was .597, Nagelkerke R^2 was .782, and McFadden R^2 was .630. These indices are not equivalent to variance explained in ordinary least-squares regression and should not be interpreted as intervention effect sizes. Nevertheless, the Nagelkerke value of .782 indicates strong model-level explanatory separation among the observed categories for this dataset. In other words, the four school-level indicators jointly contain substantial information for distinguishing numeracy tiers, while the remaining unexplained component may reflect instructional, socioeconomic, organizational, content-specific, and measurement factors not represented in Education Report.

Table 3. Pseudo R-Square

Index	Value
Cox and Snell	.597
Nagelkerke	.782
McFadden	.630

Table 4. Likelihood Ratio Tests for Individual Predictors

Effect	-2 Log Likelihood of Reduced Model	Chi-Square	df	p
Intercept	229.592	132.826	2	< .001
Reading Literacy	104.624	7.857	2	.020
Knowing / Formulating	106.526	9.759	2	.008
Applying / Employing	109.137	12.370	2	.002
Reasoning / Interpreting	103.641	6.874	2	.032

All four predictors significantly contributed to differentiating the school numeracy categories (Table 4). Reading literacy was significant ($\chi^2 = 7.857$, $p = .020$), as were Knowing/Formulating ($\chi^2 = 9.759$, $p = .008$), Applying/Employing ($\chi^2 = 12.370$, $p = .002$), and Reasoning/Interpreting ($\chi^2 = 6.874$, $p = .032$). Applying/Employing produced the largest likelihood-ratio contribution among the four predictors. This ranking should be interpreted as evidence about the comparative contribution of predictors to this fitted school-level model, not as a standardized causal effect and not as proof that increasing one indicator by a specified amount will mechanically move a school into another category.

Why Reading Literacy Matters for Numeracy

The significance of reading literacy is educationally plausible because context-rich numeracy tasks are mediated through language and representation. Before a student can mathematize a situation, the student must identify relevant information, connect quantities across sentences, interpret tables or graphics, and distinguish mathematically useful information from contextual detail. PISA-based evidence similarly shows that reading performance explains a meaningful portion of mathematics performance and can change how mathematical outcomes are interpreted (Ding & Homer, 2020). Research on multilingual science and mathematics contexts also shows that reading proficiency is closely associated with academic achievement (Stoffelsma & Spooren, 2019). The implication is not that numeracy is reducible to reading, but that inadequate access to the linguistic structure of a problem can prevent mathematical competence from becoming observable. For teachers, this supports deliberate attention to mathematical vocabulary, representations, comprehension of quantitative texts, and discussion of what information is relevant before students begin calculation.

Why Applying/Employing Provides the Largest Contribution

The largest likelihood-ratio contribution was observed for Applying/Employing. One plausible mechanism is that this phase is where mathematical knowledge must be transformed into coordinated action. Formulating a problem establishes a mathematical representation, but successful employing requires students to select strategies, execute procedures, use representations flexibly, monitor intermediate results, and connect conceptual understanding with procedural action. Stacey's (2015) account of mathematical literacy positions employing as the process in which mathematical concepts, facts,

procedures, and reasoning are actively mobilized. This interpretation is also consistent with the Mathematical Proficiency framework of Kilpatrick, Swafford, and Findell (2001): strategic competence and adaptive reasoning become visible when learners must decide what mathematics to use, justify a pathway, and adjust it when the first approach is insufficient. Thus, the current result suggests that school differences may become especially visible at the point where knowing must be converted into mathematically productive action.

The smaller likelihood-ratio contributions of Knowing/Formulating and Reasoning/Interpreting do not make those processes secondary in a developmental sense. Numeracy is cyclic and interdependent: students cannot employ mathematics appropriately without first constructing a workable representation, and a numerically correct answer remains incomplete if it is not interpreted against the original context. Research on contextual mathematical literacy shows that students often experience difficulty when tasks require sustained reasoning beyond routine procedures (Vebrarian et al., 2021). Accordingly, a productive instructional response is not to isolate Employing as a single skill, but to design tasks in which students repeatedly move through a coherent sequence of understanding the situation, structuring it mathematically, applying concepts and procedures, and evaluating the reasonableness of the solution.

Pedagogical and Epistemic Implications of the Hybrid Framework

The principal contribution of the hybrid framework is the separation of two questions that are often conflated: 'Which schools appear at risk?' and 'Which measured dimensions distinguish those school profiles?' KNN addresses the first question through predictive classification, whereas Multinomial Logistic Regression addresses the second through interpretable school-level associations. This distinction matters epistemically. A predictive model can be useful even when it does not establish causation, and an explanatory association can be educationally meaningful even when it does not guarantee out-of-sample prediction. Used together, the models can support a more disciplined form of data-informed reflection in which prediction triggers attention and explanatory indicators guide inquiry into what might need closer pedagogical examination.

For mathematics teachers, the findings suggest that numeracy development should provide repeated opportunities to read and interrogate contextual information, formulate

mathematical representations, employ concepts and procedures strategically, and interpret results in relation to real situations. Problem-based tasks, mathematical modelling, inquiry-oriented activity, and contextual numeracy problems can create environments in which these processes are coordinated rather than taught as disconnected skills. At the school and policy levels, however, the model should be treated as a screening aid rather than a high-stakes decision rule. The failure to identify the rare Intervention Required case shows why educational judgment, additional evidence, and sensitivity to minority categories remain essential. In this sense, the most useful role of Educational Data Mining is not to replace professional interpretation, but to make patterns visible early enough for professional interpretation to become more focused.

LIMITATIONS

Several limitations define the boundary within which the findings should be interpreted. First, the sample consisted of 181 elementary schools from a single metropolitan area, Makassar; therefore, the model has not yet demonstrated geographical transportability to rural settings, other provinces, or different education systems. Second, the observed class distribution was extremely imbalanced, with only one Intervention Required school. Although SMOTE can improve learning from imbalanced data, it cannot substitute for genuine empirical diversity in the rare class, and the zero recall for that category is a substantive limitation. Third, the study relied on aggregated school-level indicators. Relationships among school means cannot automatically be inferred to individual students, and doing so would risk an ecological fallacy. Fourth, the study analyzed one assessment period (2023), so it cannot establish temporal stability or describe change across assessment cycles. Fifth, the content scope was limited to the indicators available in Education Report; the analysis did not disaggregate numeracy by mathematical content domain such as number, algebra, geometry and measurement, or data and uncertainty, nor did it incorporate teacher practice, school resources, or socioeconomic characteristics. Sixth, because item-level response data were unavailable, reliability coefficients for the operational national-assessment scores could not be independently re-estimated. Finally, no instructional intervention was implemented; consequently, intervention duration and conventional treatment effect sizes are not applicable, and all findings should be interpreted as predictive and associational rather than causal evidence of instructional effectiveness.

CONCLUSION

This study demonstrates the value of integrating K-Nearest Neighbor and Multinomial Logistic Regression as complementary tools for school-level numeracy analysis. KNN achieved 90% overall accuracy on the held-out testing data and performed strongly for the Basic and Proficient categories, but its macro F1-score of 0.61 and failure to identify the single Intervention Required case show that overall accuracy alone is insufficient for judging educational usefulness under severe class imbalance. The multinomial model significantly distinguished the observed numeracy categories and showed strong model-level explanatory magnitude (Nagelkerke pseudo- $R^2 = .782$), with all four indicators contributing significantly and Applying/Employing producing the largest likelihood-ratio contribution. Together, these findings support an interpretation of numeracy as an integrated capacity in which reading access, mathematical structuring, strategic application, and contextual interpretation jointly matter.

For mathematics teachers, the practical implication is to design learning experiences that move students through the full numeracy process rather than concentrating predominantly on routine calculation. Contextual problem solving, mathematical modelling, structured discussion of quantitative texts, explicit comparison of alternative strategies, and reflection on the reasonableness of answers can strengthen the coordination between Formulating, Employing, and Interpreting. For school leaders and education authorities, the predictive-explanatory framework may be used as an early screening and reflection aid, provided that model outputs are triangulated with classroom evidence and professional judgment. Future research should conduct external validation with larger multi-regional datasets, deliberately increase representation of the lowest-performing category, examine longitudinal assessment cycles, incorporate teacher-, school-, and socioeconomic-level predictors, disaggregate numeracy by mathematical content domain, and compare the calibration and fairness of alternative machine-learning models. Such work would move Educational Data Mining in mathematics education from merely producing accurate classifications toward developing transparent, context-sensitive evidence that can responsibly inform teaching and school improvement.

ACKNOWLEDGMENTS

The authors express their gratitude to the Lembaga Pengelola Dana Pendidikan (LPDP), Ministry of Finance of the Republic of Indonesia, for financial support. Appreciation is also extended to the educators and school stakeholders in Makassar who supported the compilation and verification of the school-level Education Report data used in this study.

REFERENCES

- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. <https://doi.org/10.1613/jair.953>
- Ding, H., & Homer, M. (2020). Interpreting mathematics performance in PISA: Taking account of reading performance. *International Journal of Educational Research*, 102, 101566. <https://doi.org/10.1016/j.ijer.2020.101566>
- Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: A survey of the state of the field. *ZDM Mathematics Education*, 47(4), 531–548. <https://doi.org/10.1007/s11858-015-0708-1>
- Jain, P., & Rogers, M. (2019). Numeracy as critical thinking. *Adults Learning Mathematics: An International Journal*, 14(1), 23–33.
- Johnston, B., & Yasukawa, K. (2001). Numeracy: Negotiating the world through mathematics. In B. Atweh, H. Forgasz, & B. Nebres (Eds.), *Sociocultural research on mathematics education: An international perspective* (pp. 279–294). Lawrence Erlbaum Associates.
- Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up: Helping children learn mathematics*. National Academy Press. <https://doi.org/10.17226/9822>
- OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- Purnomo, H., Sa'dijah, C., Hidayanto, E., Sisworo, Permadi, H., & Anwar, L. (2022). Development of instrument numeracy skills test of minimum competency assessment (MCA) in Indonesia. *International Journal of Instruction*, 15(3), 635–648. <https://doi.org/10.29333/iji.2022.15335a>

- Pusmenjar. (2020a). AKM dan implikasinya pada pembelajaran. Pusat Asesmen dan Pembelajaran, Kementerian Pendidikan dan Kebudayaan.
<https://repositori.kemendikdasmen.go.id/19690/>
- Pusmenjar. (2020b). Desain pengembangan soal Asesmen Kompetensi Minimum. Pusat Asesmen dan Pembelajaran, Kementerian Pendidikan dan Kebudayaan.
- Rohim, D. C. (2021). Konsep asesmen kompetensi minimum untuk meningkatkan kemampuan literasi numerasi peserta didik sekolah dasar. *Jurnal Varidika*, 33(1), 54–62.
- Stacey, K. (2015). The international assessment of mathematical literacy: PISA 2012 framework and items. In S. J. Cho (Ed.), *Selected regular lectures from the 12th International Congress on Mathematical Education* (pp. 771–790). Springer.
https://doi.org/10.1007/978-3-319-17187-6_43
- Stoffelsma, L., & Spooren, W. (2019). The relationship between English reading proficiency and academic achievement of first-year science and mathematics students in a multilingual context. *International Journal of Science and Mathematics Education*, 17, 905–922. <https://doi.org/10.1007/s10763-018-9905-z>
- Tjahyadi, H., & Tude, K. N. L. (2025). The implementation of educational data mining in predicting students' academic achievement in mathematics at a private elementary school. *International Journal of Information and Education Technology*, 15(1), 154–163.
<https://doi.org/10.18178/ijiet.2025.15.1.2228>
- Vebrian, R., Putra, Y. Y., Saraswati, S., & Wijaya, T. T. (2021). Kemampuan penalaran matematis siswa dalam menyelesaikan soal literasi matematika kontekstual. *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, 10(4), 2602–2614.
<https://doi.org/10.24127/ajpm.v10i4.4369>
- Zhang, Y., Yun, Y., An, R., Cui, J., Dai, H., & Shang, X. (2021). Educational data mining techniques for student performance prediction: Method review and comparison analysis. *Frontiers in Psychology*, 12, 698490.
<https://doi.org/10.3389/fpsyg.2021.698490>