



PENGUATAN LITERASI SAINS MELALUI PENGENALAN METODE STATISTIKA UNTUK DATA HILANG

ENDANG S. KRESNAWATI¹, YULIA RESTI^{1*}, NOVI R. DEWI¹, IRSYADI YANI²
MARWANI²

¹Jurusan Matematika, FMIPA Universitas Sriwijaya, Inderalaya, Sumatera Selatan

²Jurusan Teknik Mesin, Fakultas Teknik Universitas Sriwijaya, Inderalaya, Sumatera Selatan

*Corresponding author: yulia_resti@mipa.unsri.ac.id

(Received: 19 Januari 2025; Accepted: 12 Juni 2025; Published on-line: 1 Juli 2025)

ABSTRAK: Kemampuan literasi sains yang baik berkontribusi secara signifikan terhadap kemajuan bangsa. Sebagai generasi penerus, sepatutnya siswa Sekolah Menengah Atas (SMA) dibekali banyak pengetahuan guna meningkatkan literasi sains yang mereka miliki termasuk dalam menangani data atau pengamatan yang hilang, mengingat saat ini merupakan era big data. Kegiatan pengabdian yang dilaksanakan ini bertujuan mengenalkan metode-metode statistika dalam mengatasi data hilang. Pengetahuan ini diharapkan dapat menguatkan literasi sains siswa SMA Kartini Palembang, yang merupakan sekolah swasta yang berada sekitar 1 km di belakang kampus Universitas Sriwijaya. Sekolah ini memiliki akreditasi B dan fokus untuk peminatan IPA. Berdasarkan pra-test dan post-test terkait pengetahuan dasar tentang metode statistika dalam menangani data hilang, hasil kegiatan menunjukkan bahwa literasi sains dan numerasi khalayak sasaran meningkat signifikan. Uji beda rata-rata kedua kelompok nilai pra-test dan post-test dengan tingkat signifikansi 5 % untuk khalayak sasaran menunjukkan bahwa kedua kelompok memiliki perbedaan yang signifikan. Hasil ini juga menginformasikan bahwa rata-rata nilai post-test secara signifikan mengalami peningkatan. Artinya kegiatan pengabdian kepada masyarakat ini memberikan implikasi yang signifikan terhadap peningkatan kemampuan literasi dan numerasi mereka.

KEY WORDS: *Data Hilang, Literasi Sains, Metode Statistika*

1. PENDAHULUAN

Rendahnya kemampuan literasi sains merupakan masalah mendasar yang memiliki dampak sangat luas bagi kemajuan bangsa. Kemampuan literasi numerasi yang rendah juga berkontribusi secara signifikan terhadap kemiskinan, pengangguran, dan kesenjangan [1]. Kemampuan literasi sains merupakan faktor penting bagi kemajuan sebuah negara dalam menjalani kehidupan di era globalisasi [2]. Namun, kemampuan literasi sains di negara-negara berkembang, khususnya Indonesia, masih sangat rendah [3]. Jika merujuk pada hasil tes PISA (*The Programme for International Student Assessment*) pada tahun 2018, Indonesia berada pada peringkat bawah, yaitu ke-71 dari 79 negara. Pada tahun 2022, Indonesia berada di peringkat ke-69 dari 80 negara yang terdaftar dalam penilaian PISA oleh *Organisation for Economic Co-operation and Development* (OECD) [4]. Skor pada peringkat terakhir ini masih berada di level bawah, sehingga Menteri Pendidikan Dasar dan Menengah (Mendikdasmen) menargetkan peningkatan skor PISA Indonesia dengan membuat pembelajaran dasar lebih mudah dan menarik [5].

Di era *big data* saat ini, diperlukan literasi tentang seluk beluk data, termasuk penanganan data atau pengamatan yang hilang. Penanganan data hilang dapat dilakukan dengan

mengimputasinya menggunakan metode-metode statistika. Data yang hilang tidak dapat diabaikan begitu saja, khususnya apabila jumlahnya cukup besar. Regresi dapat dipandang sebagai teknik analisis data dalam statistika yang memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui. Sedangkan *k-tetangga terdekat* (*k-nearest neighbors*, *k-NN*) merupakan banyaknya titik-titik pengamatan terdekat yang dipertimbangkan dalam model regresi, di mana tingkat kedekatan antara titik-titik tersebut diukur menggunakan konsep jarak, seperti Euclidean, Minkowsky, Manhattan, dan sebagainya. Konsep regresi dan *k-NN* dapat digunakan untuk mengatasi data pengamatan yang hilang dengan mengimputasinya menggunakan hasil prediksi [6]. Penelitian [7] menunjukkan bahwa konsep regresi dan *k-NN* berhasil meningkatkan kinerja metode prediksi/klasifikasi.

Sebagai sekolah swasta yang berada tidak jauh dari kampus Universitas Sriwijaya Palembang, SMA Kartini merupakan khalayak sasaran yang layak untuk kegiatan pengabdian. Adapun tujuan kegiatan ini adalah meningkatkan kemampuan literasi sains siswa SMA Kartini melalui pemahaman tentang metode-metode statistika untuk data hilang, yaitu regresi dengan *k-NN*. Topik ini sesuai dengan *Roadmap* Pengabdian pada Masyarakat FMIPA Universitas Sriwijaya Tahun 2021–2025 dengan isu strategis *Konsep Matematika dan Ilmu Pengetahuan Alam*.

2. METODE KEGIATAN

Dalam rangka mendukung program pemerintah untuk menyukseskan Gerakan Literasi Nasional, tim pengabdian mengajukan kerangka pemecahan masalah berupa kegiatan penguatan kemampuan literasi sains dengan mengenalkan metode-metode statistika untuk mengatasi data hilang, yaitu metode regresi dengan *k-NN*. Kerangka pemecahan masalah dilakukan dengan menerapkan prosedur *pretest–materi–post-test* sebanyak dua tahap. Tahap pertama untuk mengetahui kemampuan dasar khalayak sasaran tentang materi umum terkait statistika yang telah mereka pelajari di sekolah, dan tahap kedua untuk mengetahui pemahaman mereka tentang konsep regresi dan data hilang, yang merupakan pengetahuan baru dalam kegiatan ini.

Materi pertama mencakup pengertian statistik, statistika, data, informasi, ukuran pemusatan, dan ukuran penyebaran, baik untuk data tunggal maupun berkelompok. Materi kedua mencakup konsep jarak (Euclidean, Manhattan, Minkowsky), metode regresi dengan *k-NN*, dan penanganan data hilang. *Pretest* dan *post-test* berupa ujian tertulis yang terdiri dari 20 soal, dengan skor 5 untuk setiap jawaban benar dan skor total maksimal 100.

Kriteria keberhasilan kegiatan ini adalah jika minimal 70% peserta memperoleh nilai rata-rata minimal 70 pada *post-test* dan terdapat peningkatan signifikan dari *pretest*.

3. HASIL KEGIATAN

Regresi dapat dipandang sebagai teknik analisis data yang memprediksi nilai data yang tidak diketahui dengan menggunakan nilai data lain yang terkait dan diketahui. Sedangkan *k-tetangga terdekat* merupakan banyaknya titik-titik pengamatan terdekat yang dipertimbangkan dalam model regresi.

Konsep regresi dan *k-tetangga terdekat* dapat digunakan untuk mengatasi data pengamatan yang hilang dengan mengimputasinya dengan hasil prediksi [6]. Penelitian [7] menunjukkan

bahwa konsep regresi dan k-tetangga terdekat berhasil meningkatkan kinerja metode prediksi/klasifikasi.

Konsep dasar *k*-NN adalah mencari jarak terpendek antara data dalam lingkungan *k* (*jarak minimum*) yang paling dekat dengan data lengkap. Rata-rata dari data pengamatan dengan jarak terkecil menjadi nilai prediksi baru. Keuntungan metode ini adalah kesederhanaan dan efektivitasnya untuk dataset besar, serta ketahanannya terhadap *noise* [8].

Tingkat kedekatan antara titik-titik pengamatan diukur menggunakan konsep jarak, seperti jarak Euclidean, jarak Manhattan, jarak Minkowsky, dan sebagainya. Berikut konsep ketiga jarak tersebut.

3.1. Konsep Jarak Euclidean

Dalam matematika, jarak Euclidean adalah jarak garis lurus "biasa" antara dua titik dalam ruang Euclidean. Dengan jarak ini, ruang Euclidean menjadi ruang metrik. Norma yang terkait disebut norma Euclidean. Literatur lampau menyebutnya dengan metrik Pythagoras. Jarak Euclidean dihitung menggunakan persamaan (1):

$$d(x_a, x_b) = \sqrt{\sum_{j=1}^m (x_{aj} - x_{bj})^2} \quad (1)$$

Keterangan:

- $d(x_a, x_b)$: Jarak antara observasi yang mengandung *missing data* dengan yang tidak mengandung *missing data*.
- x_{aj} : Nilai dari variabel ke-*j* pada setiap observasi yang mengandung *missing data*, $j = 1, 2, \dots, m$
- x_{bj} : Nilai dari variabel lainnya pada setiap observasi yang tidak mengandung *missing data*, $j = 1, 2, \dots, m$

3.2. Konsep Jarak Manhattan

Jarak Manhattan atau bisa disebut City-Block adalah jumlah dari panjang ruas garis kedua titik pengamatan terhadap tiap sumbu dalam koordinat Kartesius. Dalam kasus data hilang jarak Manhattan dapat dipandang sebagai jarak observasi yang mengandung *missing data* dengan yang tidak mengandung *missing data* yang diasumsikan berada pada suatu sistem koordinat kartesius dengan menyusuri bagian vertikal dan horizontal tanpa pernah kembali. Nama jarak ini berasal dari tata letak jalan di pulau Manhattan yang berbentuk kisi-kisi segi empat. Jarak ini telah digunakan dalam analisis regresi sejak abad ke-18, dan saat ini umum dirujuk dengan LASSO. Interpretasi geometris dari jarak ini tercatat dari abad ke-19, terutama oleh hasil kerja Hermann Minkowski.

Secara sederhana, perhitungan jarak Manhattan dilakukan dengan mengambil nilai absolut atau nilai mutlak. Jarak Manhattan dihitung menggunakan persamaan (2):

$$d(x_a, x_b) = \sum_{j=1}^m |x_{aj} - x_{bj}| \quad (2)$$

Nilai dari jarak Manhattan bergantung pada rotasi dari sistem koordinat, namun tidak bergantung pada refleksi terhadap sumbu koordinat maupun pada translasi.

3.3. Konsep Jarak Minkowsky

Jarak Minkowsky merupakan jarak dalam ruang vektor bernorma yang dapat disebut sebagai generalisasi dari jarak Euclidean dan jarak Manhattan. Jarak Minkowsky dihitung menggunakan persamaan (3):

$$d(x_a, x_b) = \left(\sum_{j=1}^m |x_{aj} - x_{bj}|^q \right)^{1/p} \quad (3)$$

Dimana p adalah bilangan bulat. Ketika p bernilai satu disebut jarak Manhattan dan Ketika p bernilai dua disebut jarak Euclidean. Untuk limit p mendekati takhingga, didapatkan jarak Chebyshev:

$$d(x_a, x_b) = \lim_{p \rightarrow \infty} \left(\sum_{j=1}^m |x_{aj} - x_{bj}|^q \right)^{1/p} = \max_{j=1}^n |x_{aj} - x_{bj}| \quad (4)$$

Untuk p mendekati negatif takhingga, didapatkan

$$d(x_a, x_b) = \lim_{p \rightarrow -\infty} \left(\sum_{j=1}^m |x_{aj} - x_{bj}|^q \right)^{1/p} = \min_{j=1}^n |x_{aj} - x_{bj}| \quad (5)$$

Hasil pelaksanaan kegiatan tercermin pada Tabel 1 dan Gambar 1 yang mendeskripsikan nilai pra-test dan post-test peserta kegiatan PPM dengan tema pengenalan metode statistika untuk mengatasi data hilang.

Tabel 1. Resume Statistik Penilaian Peserta Kegiatan PPM

	Nilai Pra-test	Nilai Post-test
Minimum	40.00	55.00
Quartil ke-1	45.00	70.00
Rata-rata	49.80	74.20
Median	50.00	75.00
Modus	55.00	80.00
Quartil ke-3	55.00	80.00
Maksimum	60.00	85.00
Standar deviasi	6.20	7.17

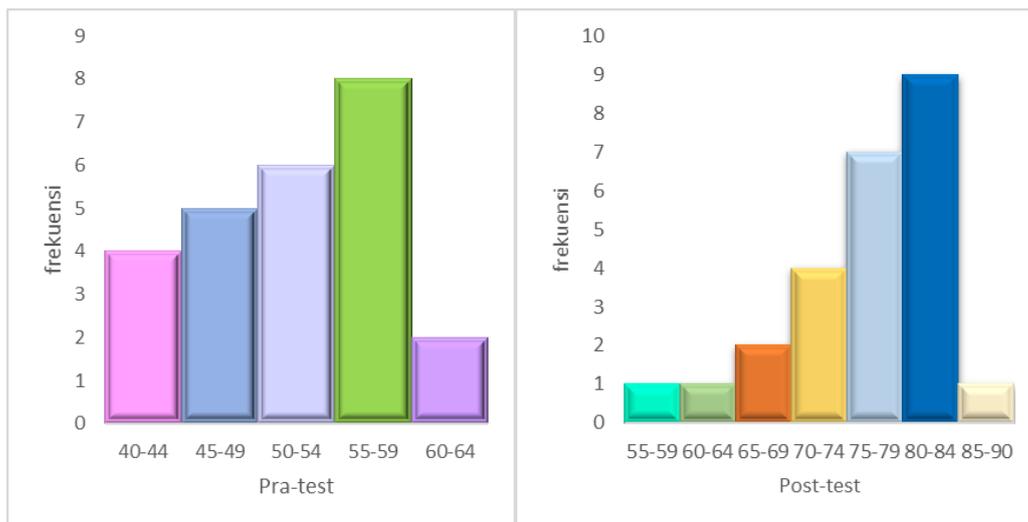
Tabel 1 menunjukkan bahwa nilai pra-test memiliki rentang nilai 40 - 60 dengan mean 49.80 dan standar deviasi 6.20. Range nilai pada hasil post-test lebih lebar yaitu 55 – 85 dengan nilai mean dan standar deviasi lebih tinggi.

Pada Gambar 1 dapat dilihat bahwa untuk nilai pra-test, frekuensi tertinggi berada pada kelompok nilai 55 - 59, diikuti berturut-turut oleh kelompok nilai 50 – 54, , 45 – 49, 40 – 44, dan 60 – 64. Untuk nilai post-test, kelompok yang memiliki frekuensi tertinggi adalah kelompok nilai 80 – 84, diikuti berturut-turut oleh kelompok 75 – 79, 70 – 74, 65– 69, dan tiga kelompok dengan frekuensi sebesar 1 yaitu kelompok nilai 55 – 59, 60 – 64, dan 85– 90.

Tabel 2 menyajikan hasil uji statistik nilai rata-rata dua kelompok pra-test dan post-test kegiatan PPM. Uji beda rata-rata kedua kelompok nilai test dengan tingkat signifikansi 5 % menunjukkan bahwa kedua kelompok memiliki perbedaan nilai yang signifikan.

Tabel 2. Hasil Pengujian hipotesis rata-rata dua kelompok

Statistik	Parameter
p-value	9.65×10^{-15}
t-stat	-12.87



Gambar 1. Histogram Nilai *Pre-test* dan *Post-test*

Hasil ini juga menginformasikan bahwa rata-rata nilai post-test secara signifikan mengalami peningkatan. Artinya kegiatan kegiatan pengabdian kepada masyarakat yang bertujuan untuk meningkatkan kemampuan literasi sains siswa SMA Kartini melalui pemahaman tentang metode-metode statistika untuk data hilang yaitu regresi dengan k-tetangga terdekat SMA Kartini Palembang memberikan dampak yang cukup memuaskan terhadap peningkatan kemampuan literasi dan numerasi mereka. Hasil rata-rata post-test yang mencapai nilai 74.20 dari semua peserta/khalayak sasaran juga menunjukkan bahwa kriteria dan indikator pencapaian tujuan yaitu lebih dari 80% khalayak sasaran berhasil mengerjakan semua soal pada post-tes, namun rata-rata hasil tes khalayak sasaran lebih dari 80 belum tercapai.

4. KESIMPULAN

Kegiatan ini telah mengenalkan penanganan data hilang menggunakan regresi dengan k-tetangga terdekat di SMA Kartini untuk mengetahui dampaknya terhadap kemampuan literasi numerasi siswa di sekolah tersebut. Hasil uji beda rata-rata kedua kelompok nilai pra-test dan post-test dengan tingkat signifikansi 5 % menunjukkan bahwa kedua kelompok memiliki perbedaan yang signifikan. Hasil ini juga menginformasikan bahwa rata-rata nilai post-test belum secara signifikan mengalami peningkatan. Peningkatan kemampuan literasi dan numerasi khalayak sasaran belum cukup memuaskan. Hasil rata-rata post-test yang mencapai nilai 74.20 dari semua peserta/khalayak sasaran juga menunjukkan bahwa kriteria dan indikator pencapaian tujuan yaitu lebih dari 80% khalayak sasaran berhasil mengerjakan semua soal pada post-tes, namun rata-rata hasil tes khalayak sasaran lebih dari 80 belum tercapai. Kegiatan sejenis perlu lebih banyak lagi diadakan di sekolah tersebut.



UCAPAN TERIMA KASIH

Ucapan terima kasih disampaikan kepada Rektor Universitas Sriwijaya melalui anggaran PNBP atas bantuan dana pada kegiatan Pengabdian kepada Masyarakat dengan Surat Keputusan Rektor Universitas Sriwijaya No. 0008/UN9.SK.LP2M.PM/2024.

DAFTAR RUJUKAN

- [1] The Conversation, "Yang harus dilakukan untuk meningkatkan tingkat literasi Indonesia," 2017. [Online]. Available: <https://theconversation.com/yang-harus-dilakukan-untuk-meningkatkan-tingkat-literasi-indonesia-83781>.
- [2] Kementerian Pendidikan dan Kebudayaan, *Panduan Gerakan Literasi Nasional*, 2017.
- [3] Tribun News, "Kemendikbud: Tingkat literasi siswa Indonesia di PISA masih rendah," 2021. [Online]. Available: <https://www.tribunnews.com/nasional/2021/03/22/kemendikbud-tingkat-literasi-siswa-indonesia-di-peringkat-pisa-masih-rendah>.
- [4] GoodStats, "Posisi Indonesia di PISA 2022," 2024. [Online]. Available: <https://goodstats.id/article/posisi-indonesia-di-pisa-2022-siapkah-untuk-2025-6RLyK>.
- [5] Suara Surabaya, "Skor PISA memprihatinkan, pemerintah diminta fokus pada pendidikan dasar," 2025. [Online]. Available: <https://www.suarasurabaya.net/kelanakota/2024/skor-pisa-memprihatinkan-pemerintah-diminta-fokus-pada-pendidikan-dasar/>.
- [6] S. Zhang et al., "A novel kNN algorithm with data-driven k parameter computation," *Pattern Recognit. Lett.*, vol. 109, pp. 44–54, 2018.
- [7] W. Chandra, B. Suprihatin, and Y. Resti, "Median-KNN Regressor-SMOTE-Tomek Links for Handling Missing and Imbalanced Data in Air Quality Prediction," *Symmetry*, vol. 15, no. 4, p. 887, 2023.
- [8] R. Alzoman and M. Alenazi, "A comparative study of distance metrics for k-nearest neighbors algorithm," *IEEE Access*, vol. 9, pp. 114975–114982, 2021.