

Pelatihan Datascience pada Pra-Pemrosesan Data untuk Siswa SMK Media Informatika - Jakarta

Datascience Training on Data Pre-Processing for Media Informatics Vocational School Students - Jakarta

Muhaimin Hasanudin^{1*}, Saruni Dwiasnati², Wawan Gunawan³

^{1,2,3}Program Teknik Informatika, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jl. Meruya Selatan
No. 1 Kembangan, Jakarta Barat, 11650 – Indonesia

*E-mail corresponding author: muhaimin.hasanudin@mercubuana.ac.id

Received: 7 April 2024; Revised: 6 Juni 2024; Accepted: 9 Oktober 2024

Abstrak. Pengabdian Kepada Masyarakat (PKM) merupakan upaya untuk melengkapi siswa dengan keterampilan yang relevan dalam teknologi informasi. Meskipun kurikulum perguruan tinggi telah menekankan pada teknologi dan ilmu komputer, kurangnya pemahaman tentang data science dan pra-pemrosesan data masih menjadi masalah. PKM ini tidak hanya mengajarkan penggunaan alat seperti Python tetapi juga mendorong pemikiran analitis dan kritis saat menganalisis data. Proses pra-pemrosesan data, termasuk normalisasi dan penanganan outlier, berdampak langsung pada input data model machine learning. PKM berfokus pada pelatihan siswa SMK Media Informatika dalam machine learning dan eksplorasi perubahan pada data masukan untuk mempengaruhi akurasi model, melalui serangkaian eksperimen menggunakan berbagai jenis dataset, program ini menganalisis bagaimana modifikasi pada data dapat mempengaruhi performa model pembelajaran mesin. Tujuan PKM adalah meningkatkan pemahaman siswa tentang pentingnya pra-pemrosesan data dalam mencapai model machine learning yang akurat. Evaluasi keberhasilan PKM melibatkan peningkatan pemahaman siswa, pengembangan keterampilan praktis, dan umpan balik positif. Pelatihan PKM berhasil meningkatkan pemahaman peserta dengan jumlah yang tidak mengerti berkurang dari 3,3% menjadi 0% setelah post-test. Persentase peserta yang mengerti materi meningkat dari 46,7% menjadi 60%.

Kata Kunci: data; datascience; pelatihan siswa; pra-pemrosesan.

Abstract. Community service is an effort to equip students with relevant skills in information technology. Although the college curriculum has emphasized technology and computer science, the lack of understanding of data science and data Pre-Processing is still an issue. Community service not only teaches the use of tools such as Python but also encourages analytical and critical thinking when analyzing data. Data Pre-Processing, including normalization and outlier handling, has a direct impact on the input data of machine learning models. The Community Service focuses on training SMK Media Informatika students in machine learning and exploring changes to the input data to affect model accuracy. Through a series of experiments with varied datasets, the Community Service analyzes the effect of changes to the data on model performance. The aim of the project was to improve students understanding of the importance of data Pre-Processing in achieving accurate machine learning models. The project's success was evaluated based on increased student understanding, practical skill development, and positive feedback. PKM training succeeded in improving participants' understanding with the number of those who did not understand decreasing from 3.3% to 0% after the post-test. The percentage of participants who understood the material increased from 46.7% to 60%.

Keywords: data; datascience; pre-processing; student training

DOI: 10.30653/jppm.v9i4.921



1. PENDAHULUAN

SMK Media Informatika Jakarta Selatan adalah sekolah menengah kejuruan pertama dalam bidang teknologi informasi (Moses, 2017). Mereka menggunakan metode pembelajaran berbasis project untuk mempersiapkan siswa menghadapi dunia kerja (Ahmad et al., 2023). Sekolah ini memiliki 10 laboratorium dan ruang praktek, termasuk 3 lab DKV, 2 lab Broadcasting, 2 lab PPLG, dan 3 lab TJKT. Setiap lab dilengkapi dengan 40 komputer, sehingga totalnya ada 400 unit komputer. Ada juga 1 Studio Siaran dan 1 Lab Bahasa. Sekolah ini telah meluluskan sekitar 5000 alumni yang sukses di dunia kerja dan usaha. Mereka juga membantu siswa yang ingin melanjutkan study ke Perguruan Tinggi Negeri (PTN) melalui seleksi nasional. SMK Media Informatika juga sering mengadakan Kelas Industri dan Job fair untuk membantu alumni menemukan pekerjaan yang sesuai dengan keahlian mereka (Nashrullah et al., 2020).

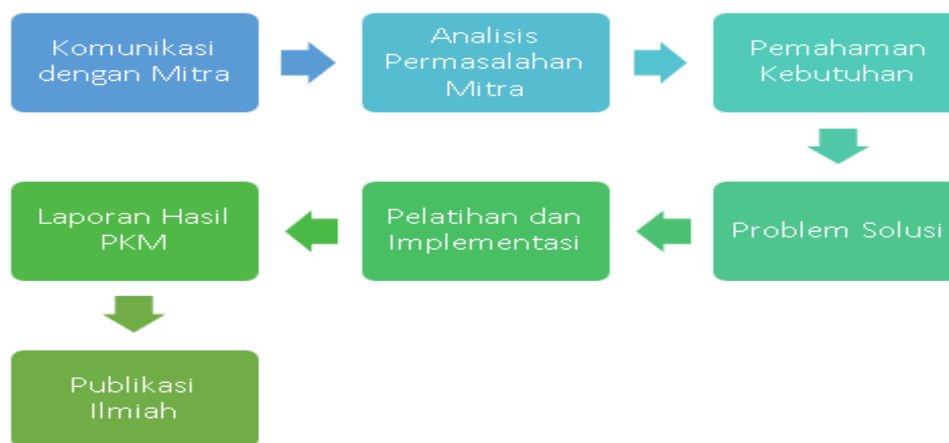
Pada era di mana kecerdasan buatan semakin mendominasi dunia teknologi, pemahaman mendalam tentang bagaimana data diproses menjadi informasi yang dapat diolah oleh model machine learning menjadi esensial. Salah satu tahapan penting dalam proses ini adalah Pra-Pemrosesan, serangkaian langkah yang dirancang untuk membersihkan, mengubah, dan mempersiapkan data sebelum disajikan ke model (Triyansyah & Fitriana, 2018). Pentingnya Pra-Pemrosesan tak dapat diabaikan. Data yang tidak terstruktur, mengandung data anomali, atau tidak lengkap dapat merugikan kinerja model dan menghasilkan prediksi yang tidak dapat diandalkan. Oleh karena itu, langkah-langkah Pra-Pemrosesan menjadi fondasi yang kuat untuk mencapai model machine learning yang akurat, efisien, dan dapat diinterpretasi (Juliansyah & Laksito, 2021; Nugraha et al., 2020). Hasil analisis dari Pra-Pemrosesan data dapat mencakup perubahan kualitas atau karakteristik data yang berdampak pada langkah selanjutnya dalam siklus pembelajaran mesin (Sudrajat & Cholid, 2023). Salah satu point setelah melakukan Pra-Pemrosesan data seperti penanganan missing values dapat menghasilkan dataset yang lebih bersih dan lengkap, menghilangkan lubang-lubang informasi yang dapat mempengaruhi analisis data (MacIver et al., 2024). Pasar kerja saat ini menuntut keterampilan dalam data science dan analitik (Prasetya et al., 2023). Dengan mengajarkan Pra-Pemrosesan agar para siswa mengetahui langkah dan cara membersihkan data mentah menjadi data jadi dan siap menjadi data model.

PKM bertujuan untuk memberikan keterampilan yang penting dalam teknologi informasi kepada siswa SMK Media Informatika (Hasanudin et al., 2023). Kurikulum perguruan tinggi perlu memperkenalkan data science dan pra-pemrosesan data. Literasi data siswa perlu ditingkatkan, termasuk kemampuan analisis, pemahaman, dan representasi data (Kristiawan & Widjaja, 2021). PKM melibatkan penggunaan tools python dan pengembangan pemikiran analitis siswa (Lionnie et al., 2022; Priambodo et al., 2020). Proses Pra-Pemrosesan melibatkan normalisasi, standarisasi, dan penanganan outliers untuk memengaruhi akurasi model machine learning (Satria et al., 2023). Metode penelitian melibatkan eksperimen dengan dataset berbagai karakteristik (Nashrullah et al., 2020). PKM bertujuan meningkatkan pemahaman dan keterampilan siswa dalam Pra-Pemrosesan untuk mencapai model machine learning yang lebih akurat.

2. METODE

Ilmu Data science berkembang dengan cepat sedangkan kurikulum sekolah terlambat mengikuti perkembangan ilmu datascience, ini bisa menyebabkan kurangnya relevansi antara sekolah dengan tuntutan industri dan kebutuhan pasar kerja (Agustiana et al., 2023). Ditambah sulitnya menemukan pengajar atau guru yang memiliki pemahaman mendalam tentang data science bisa menjadi tantangan untuk sekolah. PKM ini memberi para siswa panduan praktis untuk memilih dan menerapkan strategi Pra-Pemrosesan yang paling sesuai dengan jenis dataset yang digunakan. Adapun lokasi kegiatan PKM berada di SMK Media Informatika di Jl. Lestari II No.99, RT.4/RW.2, Petukangan Selatan, Pesanggrahan, Jakarta Selatan, Jakarta 12270. Kegiatan tersebut dilakukan secara luring atau offline sebanyak 30 siswa SMK Media Informatika. Pada tahap pelaksanaan, kegiatan PKM dilakukan pada 1 (satu) hari, yakni di Kamis, 29 Februari 2024 dimulai dari jam 09:00 hingga 15:30 WIB. terbagi menjadi 10 materi.

Langkah-langkah atau metode dalam melaksanakan solusi yang ditawarkan untuk mengatasi permasalahan mitra dari judul *Evaluasi Dampak Pra-Pemrosesan Pada Performa Model Machine Learning Untuk Siswa Smk Media Informatika*, sebagai berikut:



Gambar 1. Alur proses pelaksanaan kegiatan PKM

Proses pelaksanaan Kegiatan PkM dapat dimulai dengan melakukan komunikasi dengan mitra dan menganalisa kebutuhan dan permasalahan yang mitra hadapi. Kemudian, perumusan dari proposal PkM dibuat dengan memberikan saran atau menawarkan pelatihan datascience dan persiapan dilakukan dengan melakukan koordinasi antara mitra dan panitia yang terlibat. Setelah proses persiapan, kegiatan dilaksanakan sesuai dengan jadwal yang telah ditentukan oleh kedua belah pihak. Setelah kegiatan selesai, lalu membuat laporan dari hasil PKM yang dilakukan dan membuat artikel dari judul yang diambil.

3. HASIL DAN PEMBAHASAN

Kegiatan PKM dilakukan secara luring di SMK Media Informatika di-ikuti oleh 30 peserta siswa dengan dua tahap yakni tahap koordinasi dan tahap pelaksanaan. Pada tahap koordinasi, tim bekerja sama dengan mitra untuk membahas masalah yang sedang dihadapi, hasil dari koordinasi menunjukkan bahwa kegiatan PKM akan ditujukan kepada Siswa Sekolah SMK Media Informatika. Kegiatan juga ditetapkan untuk dilaksanakan pada hari Kamis, 29 Februari 2024. Dr. Bambang Jokonowo, SSi., MTL., Dekan Fasilkom UMB, dan Ibu Wiwin Dwi Tjiptaningsih, S.Pd., M.I.Kom., Kepala SMK Media Informatika, juga mendukung dan menyambut pelatihan data science dan *cybersecurity* ini. Sri Handayani, S.Pd., M.I.Kom., Wakil Kepala Sekolah untuk Hubungan Industri dan Hubungan Masyarakat di Sekolah Menengah Kejuruan Media Informatika Jakarta Selatan, mendorong siswa untuk meningkatkan kemampuan mereka dalam ilmu data dan keamanan *cyber*. Harapan pelatihan ini adalah agar guru dan siswa SMK lebih mampu menerapkan keterampilan data science dalam bisnis dan berbagai bidang kehidupan sehari-hari serta memperkuat mereka dalam menghadapi masalah keamanan *cyber*. Dr.Ir.Eliyani, Muhammad Rifki, S.Kom., M.Kom., Dwiki Jatikusumo, S.Kom., M.Kom., dan Bambang Sukowo adalah anggota tim dosen yang terlibat dalam kegiatan pengabdian masyarakat ini, seperti yang ditunjukkan pada gambar 2 di bawah ini :



Gambar 2. Foto bersama kegiatan PKM

Pada tahap pelaksanaan, kegiatan PKM dilakukan pada 1 (satu) hari, dimulai dari jam 09:00 hingga 15:30 WIB. Dengan demikian, kegiatan PKM terbagi menjadi 10 materi yang disampaikan dalam 1 (satu) hari. Sesi kegiatan ditunjukkan dalam tabel berikut :

Tabel 1. Program kegiatan PPM yang dilaksanakan

Hari, Tanggal	Jam	Aktivitas
Kamis 29 Februari 2024	09:00 - 09:20 WIB	<ul style="list-style-type: none"> • Pembukaan Oleh MC & Doa • Sambutan Kepala Sekolah • Sambutan Perwakilan Tim PKM
	09:20 - 09:50 WIB	Materi Pengenalan Phyton
	09:50 - 10:20 WIB	Materi Pengenalan Metode Clustering
	10:20 - 10:50 WIB	Materi Pengenalan Metode Klasifikasi
	10:50 - 11:20 WIB	Materi Pengenalan Algoritma Naive Bayes
	11:20 - 11:50 WIB	Materi Proses Pra-Pemrosesan
	11:50 - 12:20 WIB	Materi Implementasi menggunakan Naive Bayes
	12:20 - 13:00 WIB	Ishoma
	13:00 - 13:30 WIB	Materi Visualisasi menggunakan Tableau
	13:30 - 14:00 WIB	Materi <i>Cybersecurity</i>
14:00 - 14:30 WIB	Materi Pengenalan Algoritma Linear Regesi	
14:30 - 15:00 WIB	Materi Pengenalan Algoritma K-Means & Implementasi	
15:00 - 15:30 WIB	Quiz	

Pada rangka pelaksanaan kegiatan kelompok, kami mendapat giliran pada jam 11:20 - 11:50 WIB. Fasilitas kegiatan berupa proyektor, ruang lab komputer yang berada dilingkungan SMK Media Informatika. Kegiatan pengabdian masyarakat bertema Evaluasi Dampak Pra-Pemrosesan pada Performa Model Machine Learning untuk Siswa SMK Media Informatika dapat dilihat pada gambar 3 dan diuraikan sebagai berikut:

1. Kegiatan dimulai dengan memperkenalkan siswa SMK Media Informatika dengan konsep Pra-Pemrosesan dalam konteks machine learning.
2. Sebelum memulai latihan dengan model pembelajaran mesin, siswa diberi pengetahuan tentang berbagai metode Pra-Pemrosesan yang dapat diterapkan pada data.
3. Siswa melihat bagaimana masing-masing metode Pra-Pemrosesan berdampak pada kinerja pembelajaran model-mesin.

4. Siswa diberi kesempatan untuk menyampaikan hasil evaluasi mereka dan menjelaskan bagaimana temuan tersebut dapat diterapkan dalam pembelajaran di SMK Media Informatika.
5. Kegiatan ini diharapkan membantu siswa SMK Media Informatika memahami peran Pra-Pemrosesan dalam mempersiapkan data untuk analisis dengan menggunakan model pembelajaran mesin. Mereka juga akan mempelajari bagaimana memilih metode Pra-Pemrosesan yang paling sesuai dengan tujuan analisis yang ingin dicapai.



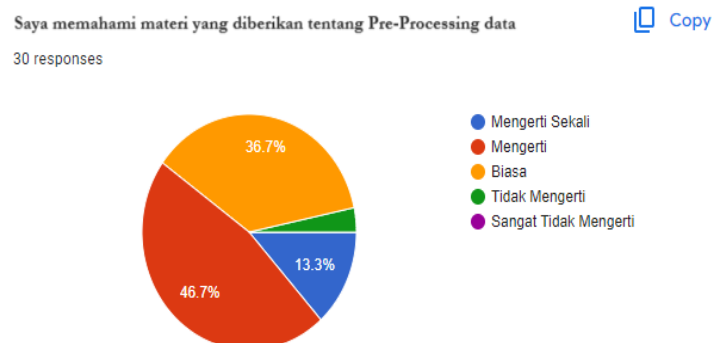
Gambar 3. Foto kegiatan PKM

Keberhasilan pelaksanaan kegiatan PKM dapat diukur melalui peningkatan pemahaman peserta mengerti apa yang telah dipaparkan oleh narasumber, pengembangan keterampilan praktis, analisis yang komprehensif, keterlibatan dan antusiasme siswa, penerapan dalam pembelajaran, serta umpan balik positif dari berbagai pihak terkait.

Tabel 2. Tolak ukur keberhasilan

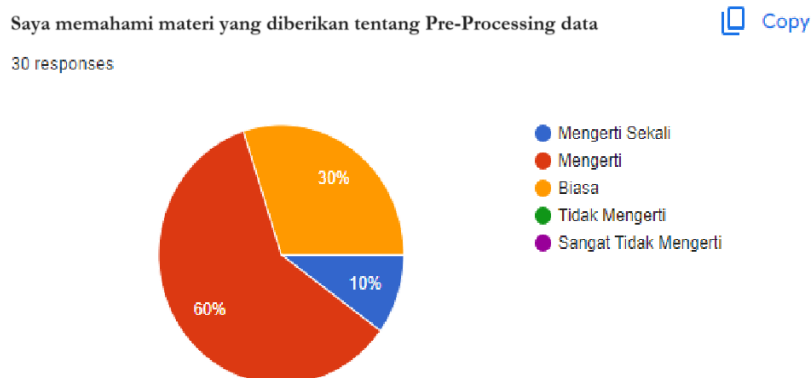
Tolak Ukur	Keterangan
Respon peserta PKM positif	Hasil evaluasi yang diisi oleh peserta PKM digunakan untuk mengukur respons mereka.

Sebelum melakukan pelatihan dilakukan pre test dengan jumlah responden 30 orang, hasil dari pre test tersebut terlihat pada gambar 4, Mengerti Sekali 13,3%, Mengerti 46.7%, Biasa 36,7%, Tidak Mengerti 3,3% dan Sangat Tidak Mengerti 0%.



Gambar 4. Pretest sebelum pelatihan

Setelah melakukan pemaparan materi dilakukan post test dengan jumlah responden 30 orang, hasil dari post test tersebut terlihat pada gambar 5, Mengerti 60%, Biasa 30% dan Mengerti Sekali 10%, terlihat untuk peserta pelatihan yang tidak mengerti tidak ada lagi.



Gambar 5. Posttest setelah pelatihan

Berdasarkan hasil pre-test dan post-test yang dilakukan terhadap 30 peserta terlihat adanya peningkatan pemahaman setelah pelatihan dilaksanakan. Pada pre-test terdapat 3,3% peserta yang tidak mengerti materi, namun pada post-test, jumlah peserta yang berada dalam kategori tidak mengerti ini berkurang hingga tidak ada. Selain itu, persentase peserta yang mengerti materi meningkat dari 46,7% pada pre-test menjadi 60% pada post-test. Meskipun persentase peserta yang mengerti sekali sedikit menurun dari 13,3% pada pre-test menjadi 10% pada post-test, tidak ada lagi peserta yang tidak mengerti, menunjukkan bahwa pelatihan berhasil meningkatkan pemahaman secara keseluruhan di kalangan peserta.

4. SIMPULAN

PKM bertujuan mengenalkan keterampilan relevan terutama di bidang ilmu datascience ke siswa di SMK Media Informatika - Jakarta. Kurikulum belum sepenuhnya mencakup aspek data science dan pra-pemrosesan data, menyebabkan literasi data siswa belum berkembang. PKM dirancang untuk mengajarkan penggunaan alat-alat seperti Python dan mendorong pemikiran analitis dan kritis siswa saat bekerja dengan data. Melalui serangkaian percobaan dengan dataset yang beragam, PKM menganalisis dampak perubahan pada data terhadap kinerja model machine learning.

UCAPAN TERIMA KASIH

Penulis yang tergabung dalam Tim Pengmas Fakultas Ilmu Komputer mengucapkan terimakasih kepada Lembaga Penelitian dan Pengabdian Kepada Masyarakat Universitas Mercu Buana Jakarta yang telah membiayai kegiatan Pengabdian masyarakat ini dengan SK No: 01-1-4/KDN-55201/002/B-SPK/II/2024 dalam mendukung pelaksanaan kegiatan PKM berjudul *Evaluasi Dampak Pra-Pemrosesan Pada Performa Model Machine Learning Untuk Siswa Smk Media Informatika*. Penulis juga mengucapkan terima kasih kepada Ibu Wiwin Dwi Tjiptaningsih, S.Pd., M.I.Kom., sebagai Kepala Sekolah di SMK Media Informatika dengan alamat Jl. Lestari II No.99, RT.4/RW.2, Petukangan Selatan, Pesanggrahan, Jakarta Selatan, Jakarta 12270.

REFERENSI

- Agustiana, A., Arisetyawan, A., & Anusaen, K. (2023). Pembelajaran Data Science di Sekolah Dasar Terhadap Kemampuan Literasi Matematika Siswa. *Jurnal Didaktika Pendidikan Dasar*, 7(3), 891–910. <https://doi.org/10.26811/didaktika.v7i3.1387>
- Ahmad, S. T., Watrionthos, R., Samala, A. D., Muskhair, M., & Dogara, G. (2023). Project-based Learning in Vocational Education: A Bibliometric Approach. *International Journal of Modern Education and Computer Science*, 15(4), 43–56. <https://doi.org/10.5815/ijmeecs.2023.04.04>
- Hasanudin, M., Rifqi, M., & Prihandi, I. (2023). Pelatihan Komputer dan Penggunaan Microsoft

- Word Pada Siswa Berkebutuhan Khusus Di Homeschooling Lantaburo Kabupaten Tangerang. *Madani : Indonesian Journal of Civil Society*, 5(2), 131–138. <https://doi.org/10.35970/madani.v5i2.1815>
- Juliansyah, S., & Laksito, A. D. (2021). Klasifikasi Citra Buah Pir Menggunakan Convolutional Neural Networks. *Jurnal Telekomunikasi Dan Komputer*, 11(1), 65. <https://doi.org/10.22441/incomtech.v11i1.10185>
- Kristiawan, K., & Widjaja, A. (2021). Perbandingan Algoritma Machine Learning dalam Menilai Sebuah Lokasi Toko Ritel. *Jurnal Teknik Informatika Dan Sistem Informasi*, 7(1), 35–46. <https://doi.org/10.28932/jutisi.v7i1.3182>
- Lionnie, R., Ramadhan, R. C., Rosyadi, A. S., Jusoh, M., & Alaydrus, M. (2022). Performance analysis of various types of surface crack detection based on image processing. *Sinergi*, 26(1), 1. <https://doi.org/10.22441/sinergi.2022.1.001>
- MacIver, C. L., Bailey, G., Laguna, P. L., Wadon, M. E., Schalkamp, A. K., Sandor, C., Jones, D. K., Tax, C. M. W., & Peall, K. J. (2024). Macro- and micro-structural insights into primary dystonia: a UK Biobank study. *Journal of Neurology*, 271(3), 1416–1427. <https://doi.org/10.1007/s00415-023-12086-2>
- Moses, K. M. (2017). The Industries Cooperation Of Information Technology Vocational High School. *Jurnal Pendidikan Sains*, 5(3), 89–95. <http://journal.um.ac.id/index.php/jps/>
- Nashrullah, F., Wibowo, S. A., & Budiman, D. G. (2020). COMPLETE Journal of Computer, Electronic, and Telecommunication Investigasi Parameter Epoch Pada Arsitektur ResNet-50 Untuk Klasifikasi Pornografi. *Journal of Computer, Electronic, and Telecommunication*, 1(1).
- Nugraha, A. F., Pristyanto, Y., & Pratama, I. (2020). Penanganan Missing Values Untuk Meningkatkan Kinerja Model Machine Learning Pada Data Telemarketing. *Pseudocode*, 7(2), 165–171. <https://doi.org/10.33369/pseudocode.7.2.165-171>
- Prasetya, M. R. A., Priyatno, A. M., & Nurhaeni. (2023). Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining. *Jurnal Informasi Dan Teknologi*, 5(2), 52–62. <https://doi.org/10.37034/jidt.v5i2.324>
- Priambodo, B., Jumaryadi, Y., & Putra, Z. P. (2020). Comparison of Local Binary Pattern and Eigenfaces for Predict Suspect Positive Drugs. *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering, BCWSP 2020*, 64–67. <https://doi.org/10.1109/BCWSP50066.2020.9249405>
- Satria, A., Badri, R. M., & Safitri, I. (2023). Prediksi Hasil Panen Tanaman Pangan Sumatera dengan Metode Machine Learning. *Digital Transformation Technology*, 3(2), 389–398. <https://doi.org/10.47709/digitech.v3i2.2852>
- Sudrajat, W., & Cholid, I. (2023). K-Nearest Neighbor (K-Nn) Untuk Penanganan Missing Value Pada Data Umkm. *Jurnal Rekayasa Sistem Informasi Dan Teknologi*, 1(2), 54–63. <https://doi.org/10.59407/jrsit.v1i2.77>
- Triyansyah, D., & Fitrihanah, D. (2018). Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing. *Jurnal Telekomunikasi Dan Komputer*, 8(3), 163. <https://doi.org/10.22441/incomtech.v8i3.4174>