

**PENERAPAN ALGORITMA DECISION TREE
UNTUK KLASIFIKASI SURAT DALAM PENGARSIPAN DIGITAL
(STUDI KASUS: SMPK RATU DAMAI)**

Petrus Ardisoma Hopi Lazar¹, Alfian Nara Weking², Bernadete Deta³

Program Studi Teknik Informatika, Fakultas Teknologi, Institut Keguruan dan Teknologi
Larantuka^{1,2,3}

Email : 1ardylazarenth01@gmail.com, 2nara.weking89@gmail.com, 3auran8150@gmail.com

Diterima: 12/9/2026; Direvisi: 1/10/2026; Diterbitkan: 6/10/2026

ABSTRAK

Pengarsipan surat yang masih dilakukan secara manual di SMPK Ratu Damai menyebabkan proses penyimpanan, pencarian, dan pengelompokan surat menjadi kurang efektif serta berisiko menimbulkan kesalahan klasifikasi. Penelitian ini bertujuan menerapkan algoritma *Decision Tree* pada sistem pengarsipan digital untuk mengklasifikasikan surat secara otomatis berdasarkan kategori arsip. Metode penelitian meliputi pengumpulan data, *preprocessing* teks yang terdiri atas *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF), serta pembentukan model *Decision Tree*. *Dataset* yang digunakan berjumlah 228 dokumen surat periode 2018–2023. Sistem dikembangkan berbasis *web* menggunakan PHP, MySQL, XAMPP, dan Visual Studio Code, kemudian dievaluasi menggunakan *Confusion Matrix* dengan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil implementasi menunjukkan bahwa sistem mampu mengotomatisasi proses klasifikasi surat, mempercepat pencarian dokumen dari 300 detik menjadi 8 detik (efisiensi 97,33%), serta mendukung pengelolaan arsip yang lebih terstruktur. Berdasarkan hasil evaluasi, model *Decision Tree* menunjukkan kinerja yang sangat baik dengan Akurasi 84,78%, Presisi 85,79%, *Recall* 86,09%, dan *F1-Score* 85,94%, sehingga layak diterapkan sebagai solusi pengarsipan digital di SMPK Ratu Damai.

Kata Kunci: *Algoritma Decision Tree, Klasifikasi Surat, Pengarsipan Digital*

ABSTRACT

The manual filing of letters at SMPK Ratu Damai makes the storage, retrieval, and grouping of letters ineffective and carries the risk of misclassification. This study aims to apply the Decision Tree algorithm to a digital archiving system to automatically classify letters based on archive categories. The research methods include data collection, text preprocessing consisting of case folding, tokenizing, stopwords removal, and stemming, TF-IDF weighting, and the development of a Decision Tree model. The dataset used consisted of 228 letter documents. The system was developed as a web-based system using PHP, MySQL, XAMPP, and Visual Studio Code. It was then evaluated using a Confusion Matrix with Accuracy, Precision, Recall, and F1-Score metrics. The implementation results showed that the system was able to automate the letter classification process, accelerate document searches from 300 seconds to 8 seconds (97.33% efficiency), and support more structured archive management. Based on the evaluation results, the Decision Tree model demonstrated strong performance with 84.78% Accuracy, 85.79% Precision, 86.09% Recall, and 85.94% F1-Score.

Keywords: *decision tree algorithm, letter classification, digital archiving*

PENDAHULUAN

Perkembangan teknologi informasi yang begitu pesat menuntut hampir semua sektor untuk beralih dari sistem manual ke sistem digital, tidak terkecuali dalam pengelolaan dokumen dan surat-menyurat di instansi pendidikan. Digitalisasi arsip kini menjadi kebutuhan penting bagi banyak institusi karena menawarkan kemudahan dalam penyimpanan, pencarian, dan pengawasan dokumen. Namun, di balik kemajuan ini, masih banyak sekolah yang belum memanfaatkan potensi teknologi secara optimal, khususnya dalam hal pengelompokan atau klasifikasi surat secara otomatis. Implementasi sistem digital yang sistematis terbukti mampu meningkatkan efisiensi waktu serta akurasi administrasi sekolah dalam menunjang operasional harian (Al-Huda et al., 2026; Hatifuddin & Junaris, 2026; Solahudin et al., 2025).

Pada SMPK Ratu Damai, pengelolaan surat masuk dan keluar sebagian besar masih dilakukan secara manual. Dokumen disimpan dalam bentuk fisik di lemari arsip atau *folder* terpisah pada komputer staf tata usaha. Metode konvensional seperti ini menimbulkan beberapa kendala utama, seperti risiko kerusakan atau kehilangan berkas fisik, keterbatasan ruang penyimpanan, waktu pencarian yang relatif lama, serta tingginya potensi kesalahan klasifikasi (*misclassification*) akibat subjektivitas petugas administrasi. Selain itu, ketergantungan pada pencatatan berbasis buku agenda yang masih bersifat konvensional memperlambat alur distribusi informasi serta menyulitkan proses audit data secara berkala (Asmarany & Prasetyawati, 2026; Hakim et al., 2026; Ishak et al., 2020; Septiana & Maulany, 2021).

Untuk mengatasi permasalahan tersebut, diterapkan algoritma *Decision Tree* dalam sistem pengarsipan digital. Algoritma *Decision Tree* dipilih karena mampu mengolah data berbasis teks dengan membentuk struktur pohon keputusan yang logis, transparan, dan mudah diinterpretasikan. Menurut penelitian terdahulu, *Decision Tree* menunjukkan performa yang sangat unggul dalam klasifikasi dokumen teks, dengan tingkat akurasi rata-rata mencapai 87–90%, lebih tinggi sekitar 4–6% dibandingkan metode *Naïve Bayes* (Kosaman & Purnomo, 2020). Implementasi model ini diharapkan dapat mengotomatisasi pengelompokan jenis surat berdasarkan kategori tertentu, sehingga proses pemilahan yang sebelumnya memakan waktu dapat diselesaikan dengan presisi yang lebih tinggi (Suminar et al., 2022).

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem pengarsipan digital berbasis *web* yang mampu mengklasifikasikan surat secara otomatis di SMPK Ratu Damai. Selain itu, penelitian ini bertujuan untuk mengevaluasi efektivitas, tingkat akurasi model, dan efisiensi waktu pemrosesannya dalam alur kerja administrasi sekolah. Pengujian dilakukan secara menyeluruh guna memastikan bahwa integrasi sistem dapat memberikan dampak operasional yang nyata bagi pihak sekolah.

Beberapa penelitian terdahulu menjadi pijakan dalam pengembangan penelitian ini. Penelitian oleh Susilowati dan Umami (2022) mengembangkan sistem informasi kearsipan berbasis *web* di SMK Tri Karya untuk mengatasi keterlambatan distribusi dan kehilangan data. Penelitian lain oleh Kuswanto et al. (2022) mengkaji aplikasi SIRADI di Universitas Negeri Semarang untuk mempercepat distribusi surat elektronik. Rohmawati dan Puspasari (2020) meneliti pengelolaan arsip dinamis di Dinas Perpustakaan dan Kearsipan Provinsi Jawa Timur, sedangkan Fauzi dan Irvansyah (2022) menyoroti pentingnya penyiapan SDM dalam transformasi digital kearsipan di SMAN 1 Takengon. Berbeda dari penelitian-penelitian tersebut, penelitian ini memfokuskan pada pengintegrasian teknik *Natural Language Processing* (NLP) dan algoritma pembelajaran mesin *Decision Tree* dengan fitur ekstraksi bobot TF-IDF untuk klasifikasi otomatis.

METODE PENELITIAN

Penelitian dilaksanakan di SMPK Ratu Damai Larantuka yang beralamat di Jl. Ki Hajar Dewantara, Kelurahan Waibalun, Kabupaten Flores Timur. Pengambilan data lapangan, observasi alur kerja arsip, serta pengujian sistem dilakukan secara langsung selama empat bulan, terhitung sejak Januari hingga April 2026. Pelaksanaan penelitian yang terjadwal ini memastikan seluruh dokumen surat dan kebutuhan sistem dapat terpetakan secara komprehensif bersama staf tata usaha sekolah. Perangkat keras yang digunakan meliputi Laptop Lenovo ThinkPad T480 dengan spesifikasi prosesor Intel Core i5-8350U, RAM 8GB, dan penyimpanan internal SSD 500GB. Perangkat lunak mencakup sistem operasi Windows 10, XAMPP (Apache web server dan basis data MySQL), Visual Studio Code sebagai *editor* kode program, Python 3.x untuk pembuatan modul klasifikasi, serta pustaka pendukung seperti *Pandas*, *Scikit-Learn*, dan *Sastrawi*. Lingkungan pengembangan ini dikonfigurasi secara lokal untuk mendukung kinerja pemrosesan teks dan integrasi basis data secara stabil.

Algoritma *Decision Tree* & Formulasinya

Algoritma *Decision Tree* membentuk pohon keputusan berdasarkan pembagian atribut secara rekursif. Pemilihan atribut terbaik pada setiap simpul (*node*) diukur berdasarkan tingkat keragaman data menggunakan nilai *Entropy* dan *Information Gain*. Rumus pengukuran *Entropy* ditunjukkan pada persamaan (1):

$$Entropy(S) = \sum_{i=1}^n -p_i \log_2(p_i)$$

Di mana S adalah himpunan sampel data, n adalah jumlah kelas, dan p_i adalah proporsi jumlah data kelas ke- i . Selanjutnya, perhitungan *Information Gain* dihitung dengan persamaan (2):

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

Di mana A adalah atribut yang diuji, $Values(A)$ adalah semua nilai yang mungkin untuk atribut A , dan S_v adalah himpunan bagian dari S yang memiliki nilai atribut v .

Tahapan Penelitian

Tahapan penelitian digambarkan dalam alur kerja sistematis yang mencakup proses pengolahan data hingga pengujian akhir:

1. **Pengumpulan Data:** Mengumpulkan 228 berkas arsip surat masuk dan keluar periode 2018–2023 dari bagian Tata Usaha SMPK Ratu Damai.
2. **Text Preprocessing:**
 - 1) **Case Folding:** Mengubah seluruh huruf menjadi huruf kecil dan menghapus tanda baca, angka, serta karakter non-alfabet.
 - 2) **Tokenizing:** Memecah kalimat dan paragraf teks menjadi kumpulan token kata tunggal.
 - 3) **Stopword Removal:** Menghilangkan kata hubung dan kata umum yang tidak memiliki bobot informasi penting.
 - 4) **Stemming:** Mentransformasikan kata berimbuhan menjadi kata dasar menggunakan algoritma Nazief-Adriani melalui pustaka *Sastrawi*.
3. **Pembobotan Fitur TF-IDF:** Menghitung bobot kata berdasarkan frekuensi kemunculan kata dalam dokumen (*Term Frequency*) dan invers frekuensi dokumen dalam korpus (*Inverse Document Frequency*).
4. **Pembentukan Model *Decision Tree*:** Membagi data menjadi 80% data latih (*training data*) dan 20% data uji (*testing data*), menghitung nilai *Entropy* dan *Information Gain*, serta menyusun pohon keputusan berbasis aturan *IF-THEN*.

5. **Evaluasi Model & Pengujian Sistem:** Mengukur kinerja klasifikasi menggunakan metrik *Confusion Matrix* serta menguji fungsionalitas sistem melalui *Black Box Testing*, pembagian hak akses (*Role-Based Access Control*), dan analisis efisiensi waktu pemrosesan.

HASIL DAN PEMBAHASAN

Hasil

Dataset penelitian terdiri atas 228 dokumen surat yang dikelompokkan ke dalam 10 kategori arsip utama sekolah. Tahapan *text preprocessing* diterapkan pada seluruh isi dan perihal surat untuk membersihkan kata-kata yang tidak relevan sehingga diperoleh bentuk kata dasar yang seragam. Contoh hasil tahapan *text preprocessing* pada cuplikan perihal surat ditunjukkan pada Tabel 1.

Tabel 1. Contoh Tahapan Text Preprocessing Surat

Tahapan	Hasil Teks
Dokumen Asli	Dengan Hormat, Bersama surat ini kami mengajukan Permohonan Bantuan Dana Kegiatan Sekolah Tahun 2026.
Case Folding	dengan hormat bersama surat ini kami mengajukan permohonan bantuan dana kegiatan sekolah tahun 2026
Tokenizing	[dengan, hormat, bersama, surat, ini, kami, mengajukan, permohonan, bantuan, dana, kegiatan, sekolah]
Stopword Removal	[hormat, surat, mengajukan, permohonan, bantuan, dana, kegiatan, sekolah]
Stemming	[hormat, surat, aju, mohon, bantu, dana, giat, sekolah]

Sistem pengarsipan digital diimplementasikan berbasis *web* dengan menerapkan skema *Role-Based Access Control* (RBAC) yang membagi hak akses ke dalam tiga tingkatan pengguna. Tingkat pertama adalah *Administrator* yang memiliki wewenang penuh untuk mengelola akun pengguna, data latih, data arsip surat, serta mengeksekusi model klasifikasi. Tingkat kedua adalah Staf Tata Usaha yang bertugas mencatat dan mengunggah surat masuk dan keluar secara harian. Tingkat ketiga adalah Kepala Sekolah yang memiliki hak akses *read-only* guna meninjau rekapitulasi statistik arsip dan laporan dokumen secara berkala.

Pengujian model klasifikasi dilakukan terhadap 46 dokumen surat uji (20% dari total 228 data). Kinerja model dievaluasi melalui matriks *Confusion Matrix* multikelas untuk mengetahui nilai *Precision*, *Recall*, dan *F1-Score* pada setiap kategori surat. Rekapitulasi hasil pengujian model *Decision Tree* dirangkum pada Tabel 2.

Tabel 2. Rekapitulasi Metrik Evaluasi Model Decision Tree per Kategori

Kategori Surat	Precision (%)	Recall (%)	F1-Score (%)
Surat Undangan	87.50%	87.50%	87.50%
Surat Pemberitahuan	80.00%	80.00%	80.00%
Surat Laporan	75.00%	85.71%	80.00%
Surat Permohonan	88.89%	80.00%	84.21%
Surat Instruksi	83.33%	83.33%	83.33%
Surat Keuangan	100.00%	100.00%	100.00%
Rata-rata (Macro Average)	85.79%	86.09%	85.94%

Dari total 46 data uji yang dianalisis, sistem berhasil memprediksi 39 dokumen dengan tepat dan mengalami kesalahan klasifikasi pada 7 dokumen. Berdasarkan perbandingan nilai prediksi benar terhadap total data uji, nilai Akurasi keseluruhan (*Overall Accuracy*) model *Decision Tree* diperoleh sebesar 84,78%.

Pengujian efisiensi waktu dilakukan untuk membandingkan durasi pengarsipan dokumen surat secara manual dengan pengarsipan digital otomatis. Perbandingan operasional dan hasil pencatatan durasi disajikan pada Tabel 3.

Tabel 3. Hasil Pengujian Efisiensi Waktu Pengarsipan

Metode Pengarsipan	Rata-rata Waktu Proses	Keterangan Operasional
Pengarsipan Manual	300 detik (5 menit)	Membaca dokumen fisik, mencatat buku agenda manual, mencari map arsip, menyimpan fisik di lemari
Sistem Digital (<i>Decision Tree</i>)	8 detik	Unggah digital, <i>auto preprocessing</i> , <i>auto classification</i> , dan penyimpanan basis data otomatis

Peningkatan efisiensi waktu dihitung menggunakan formula:

$$Efisiensi = \frac{300 - 8}{300} \times 100\% = 97,33\%$$

Hasil perhitungan tersebut membuktikan bahwa adopsi sistem digital mampu mereduksi durasi kerja administrasi secara signifikan di tingkat operasional.

Pembahasan

Berdasarkan metrik evaluasi pada Tabel 2, performa klasifikasi memperlihatkan perbedaan karakteristik yang cukup mencolok di antara kategori dokumen. Kategori Surat Keuangan menghasilkan performa paling sempurna dengan capaian *Precision*, *Recall*, dan *F1-Score* sebesar 100,00%. Hal ini disebabkan oleh keberadaan kata kunci yang sangat spesifik dan konsisten pada korpus dokumen keuangan, seperti kata "anggaran", "kuitansi", "biaya", "honor", dan "dana", sehingga pohon keputusan dapat membentuk aturan pemisah (*splitting rules*) dengan nilai *Information Gain* yang tinggi tanpa ambiguitas antarkelas. Sebaliknya, kategori dengan keragaman kosakata yang lebih luas cenderung menghasilkan metrik yang lebih rendah akibat adanya tumpang tindih fitur antarlabel (Antoh et al., 2025; Hancock et al., 2024; Khanday et al., 2020; Muñoz & Iglesias, 2022).

Sebaliknya, metrik terendah ditemukan pada kategori Surat Laporan yang mencatatkan nilai *Precision* sebesar 75,00% serta Surat Pemberitahuan dengan *Precision* 80,00%. Penurunan performa pada kedua kelas ini dipicu oleh adanya tumpang tindih leksikal (*lexical overlap*) pada isi teks surat, di mana kata-kata umum seperti "kegiatan", "laksana", "hasil", dan "waktu" sering muncul bersamaan baik dalam surat laporan pertanggungjawaban maupun surat pemberitahuan kegiatan. Kemunculan fitur kata yang serupa pada lebih dari satu kelas mengakibatkan model *Decision Tree* mengalami salah klasifikasi (*misclassification*) pada 7 dokumen uji karena bobot TF-IDF kata-kata tersebut tidak cukup dominan untuk memisahkan cabang keputusan secara tegas. Analisis lebih lanjut menunjukkan bahwa peningkatan jumlah fitur atau penggunaan teknik normalisasi yang berbeda dapat mengurangi tingkat ambiguitas ini (Eykens et al., 2020; Rianto et al., 2026; Rohmatun & Baita, 2025; Supono & Suprayogi, 2021; Zamsuri et al., 2024).

Capaian akurasi keseluruhan sebesar 84,78% pada penelitian ini sejalan dengan temuan Kosaman dan Purnomo (2020) yang mencatat bahwa algoritma *Decision Tree* memiliki performa klasifikasi teks berkisar antara 87–90%. Sedikit lebih rendahnya akurasi pada penelitian ini disebabkan oleh variasi korpus dokumen yang lebih heterogen pada surat administrasi sekolah serta keterbatasan jumlah data latih pada kategori-kategori minoritas. Meskipun demikian, akurasi 84,78% ini sudah terbukti andal untuk diterapkan di lingkungan sekolah, terlebih dengan efisiensi waktu pemrosesan sebesar 97,33% yang mampu mengeliminasi hambatan penumpukan berkas fisik di SMPK Ratu Damai. Implementasi model

ini menunjukkan bahwa otomasi klasifikasi berbasis pohon keputusan mampu meningkatkan efisiensi operasional secara signifikan dibandingkan dengan proses penyortiran manual yang sebelumnya memakan waktu (Bella & Fatah, 2025; Maulana et al., 2024; Musyarofah et al., 2024; Paramartha et al., 2020; Rahimi et al., 2025).

KESIMPULAN

Penerapan algoritma *Decision Tree* yang diintegrasikan dengan teknik *text preprocessing* dan pembobotan fitur TF-IDF pada sistem pengarsipan digital berbasis *web* terbukti mampu mengatasi permasalahan pengelolaan arsip di SMPK Ratu Damai. Sistem yang dibangun berhasil mengotomatisasi proses klasifikasi surat ke dalam kategori yang tepat dengan kinerja model yang sangat baik, ditunjukkan oleh capaian Akurasi sebesar 84,78%, Presisi 85,79%, *Recall* 86,09%, dan *F1-Score* 85,94%. Selain itu, implementasi sistem mampu memangkas waktu kerja administrasi pengarsipan secara drastis dari 300 detik menjadi hanya 8 detik per dokumen atau mencapai tingkat efisiensi waktu sebesar 97,33%.

Penelitian ini masih memiliki keterbatasan pada ukuran korpus data latih yang terbatas pada 228 dokumen serta adanya tumpang tindih kosakata pada kategori dokumen tertentu yang memicu kesalahan klasifikasi. Oleh karena itu, penelitian lanjutan disarankan untuk memperluas jumlah dataset surat, menerapkan metode seleksi fitur tambahan seperti *Chi-Square* atau *Information Gain Thresholding*, serta mengeksplorasi penggunaan algoritma *ensemble* seperti *Random Forest* atau algoritma *deep learning* untuk meningkatkan ketahanan klasifikasi terhadap teks dengan kemiripan kata yang tinggi.

DAFTAR PUSTAKA

- Al-Huda, M. S., Ali, M., Cholid, N., & Ekaningrum, I. R. (2026). *Total quality management dalam sistem manajemen berbasis madrasah di SMP. LEARNING: Jurnal Inovasi Penelitian Pendidikan dan Pembelajaran*, 6(1), 575–586. <https://doi.org/10.51878/learning.v6i1.9242>
- Antoh, S., Herteno, R., Budiman, I., Kartini, D., & Mazdadi, M. I. (2025). Prediksi *churn* pelanggan telekomunikasi dengan optimalisasi seleksi fitur dan *tuning hyperparameter* pada algoritma klasifikasi C4.5. *Jurnal Sistem Informasi Bisnis*, 15(1), 60–67. <https://doi.org/10.14710/vol15iss1pp60-67>
- Asmarany, S. A., & Prasetyawati, A. S. W. (2026). Analisis sistem informasi *monitoring* surat masuk dan surat keluar pada KUA Kecamatan Tarub. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 10(1), 829–835. <https://doi.org/10.36040/jati.v10i1.16876>
- Bella, S. L., & Fatah, Z. (2025). Klasifikasi status akademik mahasiswa menggunakan *decision tree*. *Jurnal Mahasiswa Teknik Informatika*, 4(2), 141–146. <https://doi.org/10.35473/jamastika.v4i2.4523>
- Eykens, J., Guns, R., & Engels, T. (2020). Fine-grained classification of social science journal articles using textual data: A comparison of supervised machine learning approaches. *Quantitative Science Studies*, 2(1), 89–110. https://doi.org/10.1162/qss_a_00106
- Fauzi, M., & Irvansyah, A. (2022). Transformasi digital pada sistem kearsipan di SMAN 1 Takengon Kabupaten Aceh Tengah. *Jurnal Manajemen Pendidikan dan Informasi*, 4(1), 33–41. <https://doi.org/10.51878/jmpi.v4i1.1892>
- Hakim, L., Kristanto, S. P., & Wibowo, G. H. (2026). Penguatan layanan informasi dan laporan masyarakat di Desa Kalirejo melalui implementasi *chatbot* WhatsApp terintegrasi

- sistem *dashboard*. *Archive: Jurnal Pengabdian Kepada Masyarakat*, 5(2), 446–460. <https://doi.org/10.55506/arch.v5i2.272>
- Hancock, J., Wang, H., Khoshgoftaar, T. M., & Liang, Q. (2024). Data reduction techniques for highly imbalanced medicare big data. *Journal of Big Data*, 11(1), Article 34. <https://doi.org/10.1186/s40537-023-00869-3>
- Hatifuiddin, M. I., & Junaris, I. (2026). Implementasi pengelolaan sistem informasi manajemen lembaga pendidikan berbasis *website* di MI Plus Walisongo Trenggalek. *COMMUNITY: Jurnal Pengabdian Kepada Masyarakat*, 6(1), 164–174. <https://doi.org/10.51878/community.v6i1.8952>
- Ishak, R., Setiaji, Akbar, F., & Safudin, M. (2020). Rancang bangun sistem informasi surat masuk dan surat keluar berbasis *web* menggunakan metode *waterfall*. *Jurnal Indonesia Sosial Teknologi*, 1(3), 198–209. <https://doi.org/10.36418/jist.v1i3.33>
- Khanday, A. M. U. D., Rabani, S. T., Khan, Q. R., Rouf, N., & Din, M. M. U. (2020). Machine learning based approaches for detecting COVID-19 using clinical text data. *International Journal of Information Technology*, 12(3), 731–739. <https://doi.org/10.1007/s41870-020-00495-9>
- Kosaman, J., & Purnomo, H. (2020). Perbandingan kinerja *decision tree* dan *naïve bayes* dalam klasifikasi dokumen teks. *Jurnal Rekayasa Sistem dan Teknologi Informasi*, 8(3), 201–209. <https://doi.org/10.29207/resti.v8i3.2140>
- Kuswanto, A., Ungu, R. B. M., Rahmahwati, W. D., & Rahmawati, F. D. (2022). Manajemen surat masuk dan surat keluar di Universitas Negeri Semarang melalui SIRADI. *Jurnal Kearsipan dan Administrasi Perkantoran*, 7(2), 105–115. <https://doi.org/10.15294/jkap.v7i2.4182>
- Maulana, R., Hassolthine, C. R., & Saputra, M. I. (2024). Implementasi algoritma *decision tree* dalam optimasi penilaian kinerja operasional. *Journal of Information System, Applied, Management, Accounting and Research*, 8(1), 161–170. <https://doi.org/10.52362/jisamar.v8i1.1432>
- Muñoz, S., & Iglesias, C. Á. (2022). A text classification approach to detect psychological stress combining a lexicon-based feature framework with distributional representations. *Information Processing & Management*, 59(5), Article 103011. <https://doi.org/10.1016/j.ipm.2022.103011>
- Musyarofah, M., Martanto, M., & Hayati, U. (2024). Klasifikasi kompetensi siswa di SMK Basuraga menggunakan algoritma *decision tree*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 466–472. <https://doi.org/10.36040/jati.v8i1.8354>
- Paramartha, A. A. G. Y., Marti, N. W., & Aryanto, K. Y. E. (2020). Comparison of classification model and annotation method for Undiksha's official documents. *Journal of Physics: Conference Series*, 1516(1), Article 012026. <https://doi.org/10.1088/1742-6596/1516/1/012026>
- Rahimi, H., Defit, S., & Veri, J. (2025). Implementasi *decision tree* dalam pengambilan keputusan untuk pemberian beasiswa. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4175–4180. <https://doi.org/10.36040/jati.v9i3.13580>
- Rianto, R., Humanika, E. S., & Untoro, I. H. T. (2026). Enhancing SVM-based classification performance on Indonesian sentences through TF-IDF and directional augmentation. *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, 10(1), 22–35. <https://doi.org/10.29407/intensif.v10i1.25179>
- Rohmatun, L., & Baita, A. (2025). Machine learning-based sentiment analysis on Twitter (X): A case study of the “Kabur Aja Dulu” issue using SVM. *Journal of Applied*



Informatics and Computing, 9(4), 1972–1983.
<https://doi.org/10.30871/jaic.v9i4.9991>

- Rohmawati, N., & Puspasari, D. (2020). Pengelolaan arsip berbasis aplikasi surat di Dinas Perpustakaan dan Kearsipan Provinsi Jawa Timur. *Jurnal Administrasi Perkantoran*, 8(1), 50–61. <https://doi.org/10.26740/joa.v8n1.p50-61>
- Septiana, T. D., & Maulany, R. (2021). Pengembangan manajemen data dan informasi menggunakan analisis *soft system methodology* di Universitas Advent Indonesia. *TeIKa: Jurnal Teknologi Informasi dan Komunikasi*, 11(1), 1–13. <https://doi.org/10.36342/teika.v11i1.2473>
- Solahudin, M., Jayadi, J., & Irfan, J. A. (2025). Implementation of digital-based school management systems to improve administrative efficiency in junior high schools in East Kalimantan Province. *JiIP: Jurnal Ilmiah Ilmu Pendidikan*, 8(9), 10960–10967. <https://doi.org/10.54371/jiip.v8i9.9369>
- Suminar, S., Mareta, Y. A., Kurniawan, M. A., & Wijayanti, D. G. S. (2022). Klasifikasi persuratan dengan menggunakan algoritma *decision tree* di Lembaga Penelitian dan Pengabdian Kepada Masyarakat. *Jurnal Penelitian Pendidikan*, 39(1), 32–42. <https://doi.org/10.15294/jpp.v39i1.37192>
- Supono, R. A., & Suprayogi, M. A. (2021). Perbandingan metode TF-ABS dan TF-IDF pada klasifikasi teks *helpdesk* menggunakan *K-Nearest Neighbor*. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(5), 911–918. <https://doi.org/10.29207/resti.v5i5.3403>
- Susilowati, I., & Umami, I. (2022). Sistem informasi kearsipan surat masuk dan surat keluar pada SMK Tri Karya. *Jurnal Teknologi dan Sistem Informasi*, 3(2), 90–98. <https://doi.org/10.31004/jtsi.v3i2.842>
- Zamsuri, A., Defit, S., & Nurcahyo, G. W. (2024). Development and comparison of multiple emotion classification models in Indonesia text using machine learning. *Journal of Advances in Information Technology*, 15(4), 519–531. <https://doi.org/10.12720/jait.15.4.519-531>