



Penerapan Data Mining untuk Kelulusan Al-Barqi di Pesantren Salafiyah Syafi'iyah Menggunakan Metode Random Forest

Zaehol Fatah^{1*}, Faza Qori Aina²

¹Sistem Informasi, Sains dan Teknologi, Universitas Ibrahimy, Indonesia

²Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy, Indonesia

*Penulis Korespondensi: zaeholfatah@gmail.com

Abstract. Application of data mining methods using Random Forest algorithm to predict student graduation in the Al-Barqi learning program at Pesantren Salafiyah Syafi'iyah. The background of this research is the continued use of manual evaluation systems that have the potential to create subjectivity and lack transparency. The data used comes from student evaluation results, covering aspects of fasahah, tilawah, numerical scores, and discipline. The research process was conducted through stages of data collection, preprocessing, modeling, and implementation using RapidMiner Studio software. The research results show that the Random Forest algorithm is able to provide predictions with a good level of accuracy, as well as identify the score factor as the most dominant variable in determining graduation, followed by aspects of Quranic reading and discipline. These findings confirm that the use of data mining can improve accuracy, objectivity, and transparency in student assessment systems, while supporting digital transformation in pesantren environments. Thus, this research provides data-based recommendations to improve the quality of learning and evaluation systems in pesantren.

Keywords: Al-Barqi; Data Mining; Pesantren; Random Forest; Rapid Miner.

Abstrak. Penerapan metode data mining menggunakan algoritma Random Forest untuk memprediksi kelulusan santri pada program pembelajaran Al-Barqi di Pesantren Salafiyah Syafi'iyah. Latar belakang penelitian ini adalah masih digunakannya sistem evaluasi manual yang berpotensi menimbulkan subjektivitas dan kurang transparan. Data yang digunakan berasal dari hasil evaluasi santri, meliputi aspek fasahah, tilawah, nilai numerik, serta kedisiplinan. Proses penelitian dilakukan melalui tahapan pengumpulan data, processing, pemodelan, dan implementasi dengan menggunakan perangkat lunak RapidMiner Studio. Hasil penelitian memperlihatkan bahwa algoritma Random Forest mampu menghasilkan prediksi dengan tingkat akurasi yang tinggi, serta mengidentifikasi faktor nilai sebagai variabel paling dominan dalam menentukan kelulusan, diikuti oleh aspek bacaan Al-Qur'an dan kedisiplinan. Temuan ini menegaskan bahwa penggunaan data mining dapat meningkatkan akurasi, objektivitas, dan transparansi dalam sistem penilaian santri, sekaligus mendukung transformasi digital di lingkungan pesantren. Dengan demikian, penelitian ini memberikan rekomendasi berbasis data untuk meningkatkan kualitas pembelajaran dan sistem evaluasi di pesantren.

Kata kunci: Al-Barqi; Data Mining; Pesantren; Random Forest; Rapid Miner.

1. LATAR BELAKANG

Salah satu bentuk implementasi pendidikan Al-Qur'an di lingkungan pesantren adalah melalui Taman Pendidikan Al-Qur'an (TPQ). Di Pesantren Salafiyah Syafi'iyah, TPQ ini dikenal dengan nama Al-Barqi, yang menjadi wadah utama bagi santri terutama anak-anak usia dini hingga remaja untuk mempelajari baca-tulis Al-Qur'an, tajwid, hafalan (tahfizh), serta nilai-nilai akhlak Islami secara intensif (Masnawati & Fitria, 2024).

Kelulusan dalam program Al-Barqi tidak hanya diukur dari kemampuan teknis membaca Al-Qur'an, tetapi juga mencakup aspek kedisiplinan, kehadiran, hafalan, pemahaman tajwid, serta partisipasi dalam kegiatan keagamaan sehari-hari. Proses evaluasi kelulusan santri Al-Barqi selama ini masih dilakukan secara manual atau semi otomatis berdasarkan penilaian subjektif ustadz/ustadzah, sehingga berpotensi menimbulkan ketidakkonsistenan dan kurangnya transparansi dalam pengambilan keputusan (Firdausiyah Luluk et al., 2022).

Dalam era digitalisasi pendidikan, pemanfaatan teknologi informasi khususnya data mining, menjadi solusi strategis untuk menganalisis pola dan hubungan antar variabel yang memengaruhi kelulusan santri. Data mining memungkinkan pengelola pesantren untuk menggali informasi berharga dari data historis santri, sehingga dapat digunakan untuk memprediksi dan meningkatkan kualitas output pendidikan. Salah satu metode dalam data mining yang telah terbukti efektif dan andal. menangani permasalahan klasifikasi adalah Random Forest. Metode ini mampu memeberikan model prediksi yang akurat, Konsisten serta mampu bertahan terhadap gangguan atau ketidakteraturan data, serta mampu menangani berbagai jenis variabel sekaligus(Sobari et al., 2025). Hasil dari penelitian ini adalah berbasis data kepada pengelola pesantren dalam merancang intervensi pembelajaran yang tepat sasaran, meningkatkan akuntabilitas sistem evaluasi, serta mendorong transformasi digital di lingkungan pendidikan pesantren yang selama ini berbasis tradisi lisan dan manual. Dengan demikian, integrasi teknologi dan nilai-nilai kepesantrenan dapat berjalan beriringan demi mencetak generasi Qur'ani yang unggul dan berkarakter.penelitian.

2. KAJIAN TEORITIS

Data Mining

Data mining adalah proses mengidentifikasi pola serta informasi berharga dari sekumpulan data. berukuran besar dengan memanfaatkan berbagai teknik statistik dan machine learning untuk membantu pengambilan keputusan. Teknik utamanya meliputi classification, clustering, dan regression yang diterapkan di berbagai sektor seperti retail, keuangan, dan healthcare. Prosesnya dimulai dari pengumpulan data, pemilihan algoritma, pembangunan model, hingga penerapan praktis, menjadikannya alat penting untuk mengubah data mentah menjadi wawasan yang berguna bagi organisasi(Oon Wira Yuda et al., 2022).

Random Forest

Random forest adalah teknik pembelajaran mesin berbasis ensemble learning yang digunakan untuk menangani berbagai permasalahan seperti klasifikasi, regresi, dan tugas-tugas lainnya. Teknik ini bekerja dengan membangun banyak pohon keputusan (decision trees) secara acak, lalu mengombinasikan seluruh hasil prediksi dari pohon-pohon tersebut menggunakan sistem pemungutan suara mayoritas pada permasalahan klasifikasi atau menggunakan perhitungan rata-rata pada permasalahan regresi. Prinsip kerja algoritma ini didasarkan pada pembangunan multiple pohon keputusan yang kemudian diintegrasikan untuk menghasilkan prediksi yang lebih konsisten dan tepat. Kumpulan pohon keputusan yang membentuk "hutan" dalam algoritma ini umumnya dilatih menggunakan teknik bagging

(bootstrap aggregating), yang bertujuan untuk menciptakan model pembelajaran mesin dengan performa yang lebih optimal (Abdul Khalim et al., 2023).

Rapid Miner

RapidMiner adalah sebuah perangkat lunak (software) berbasis data science yang digunakan untuk melakukan data mining, machine learning, text mining, dan analisis prediktif. RapidMiner membantu pengguna mengolah data dalam jumlah besar untuk menemukan pola, tren, maupun pengetahuan baru, tanpa harus banyak menulis kode (karena menggunakan sistem drag-and-drop) (Vibrianti et al., 2024).

Pembelajaran Al-Qur'an

Pendidikan Al-Qur'an adalah suatu kegiatan pembelajaran yang dirancang untuk membantu umat Islam dalam mempelajari, menghafalkan, dan mengimplementasikan nilai-nilai yang terdapat di dalam Al-Qur'an sebagai panduan kehidupan. Pembelajaran ini tidak hanya berfokus pada kemampuan membaca Al-Qur'an dengan benar sesuai kaidah tajwid agar terjaga keautentikan dan maknanya, tetapi juga meliputi pemahaman terhadap makna ayat-ayat serta pengaplikasian prinsip-prinsip Al-Qur'an dalam aktivitas keseharian. Keberhasilan pembelajaran Al-Barqi dievaluasi melalui kelancaran bacaan, ketepatan tajwid, jumlah hafalan, dan kualitas spiritual santri (Fitriana et al., 2025).

Pesantren

Pesantren adalah lembaga yang bersifat tradisional dan berkembang di Indonesia untuk pembelajaran ilmu agama dan pembentukan karakter santri. Pesantren dipimpin Kyai dengan lima elemen yaitu: Kyai, santri, pondok, masjid, dan kitab kuning. Santri tinggal di pondok untuk mendalami ilmu agama seperti fiqh, hadits, dan tafsir sambil membentuk akhlak dan kemandirian. Pesantren dibedakan menjadi salafiyah (ilmu agama tradisional), khalafiyah (integrasi pendidikan umum), dan campuran, berperan penting dalam pembelajaran tradisi Islam dan pembentukan ulama (Sihabuddin et al., 2024).

3. METODE PENELITIAN

Metode Pengumpulan Data

Data yang digunakan untuk penelitian ini berasal dari sumber yang telah ditentukan oleh peneliti Pondok Pesantren Salafiyah Syafi'iyah, khususnya dari program pembelajaran Al-Barqi. Data yang dikumpulkan meliputi informasi terkait hasil evaluasi santri seperti kelancaran bacaan, ketepatan tajwid, jumlah hafalan, serta aspek kedisiplinan dan penilaian spiritual. Selain itu, data pendukung berupa catatan kehadiran, nilai ujian, dan hasil observasi guru juga dijadikan sebagai variabel penelitian. Melalui wawancara dengan ustadz/ustadzah

pengampu untuk memastikan validitas informasi(Warahmah et al., 2023). Data yang terkumpul kemudian diolah menjadi dataset yang siap digunakan dalam proses data mining dengan algoritma Random Forest.

Pra-pemrosesan Data (Preprocessing)

Tahap pra-pemrosesan data dilakukan untuk menyiapkan dataset agar sesuai dengan kebutuhan analisis data mining. Di antaranya yaitu:

1. Pembersihan data (data cleaning), yaitu mengatasi data yang tidak lengkap, duplikat, atau tidak konsisten agar tidak memengaruhi hasil analisis.
2. Transformasi data, yaitu mengubah data ke dalam format yang sesuai, misalnya mengonversi nilai kualitatif (seperti predikat penilaian) menjadi nilai kuantitatif.
3. Normalisasi data, yaitu menyeragamkan skala data agar setiap atribut memiliki bobot yang seimbang dalam proses perhitungan.
4. Seleksi atribut (feature selection), yaitu menentukan variabel yang relevan dan berpengaruh terhadap kelulusan santri, seperti bacaan, tajwid, hafalan, dan kedisiplinan.

Hasil dari tahap pra-pemrosesan adalah dataset yang bersih, konsisten, dan terstruktur sehingga siap digunakan pada tahap penerapan algoritma Random Forest(Magriza, 2022).

Pemodelan Data

Tahap pemodelan data dilakukan dengan menerapkan Algoritma Random Forest digunakan dalam proses pembentukan model prediksi kelulusan santri pada pembelajaran Al-Barqi. Proses pemodelan diawali dengan membagi dataset Data tersebut Data dibedakan menjadi dua jenis, algoritma Random Forest membentuk sejumlah pohon keputusan (decision trees) dari data latih, kemudian melakukan proses voting untuk menentukan hasil prediksi kelulusan santri. Model yang dihasilkan pada tahap ini kemudian digunakan sebagai dasar dalam melakukan evaluasi dan analisis hasil untuk mengetahui faktor-faktor yang berpengaruh terhadap kelulusan(Diantika et al., 2024).

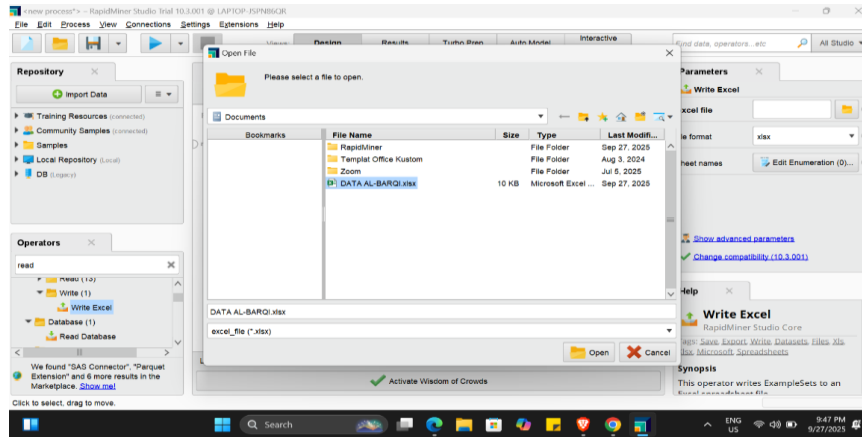
Implementasi

Tahap implementasi adalah penerapan model Random Forest pada dataset santri Model yang telah dilatih dengan data latih diuji menggunakan data uji untuk memprediksi kelulusan santri. Proses ini dilakukan dengan RapidMiner, mulai dari input data, penerapan algoritma, hingga menghasilkan output berupa klasifikasi santri dalam kategori lulus atau belum lulus. Hasil implementasi menjadi dasar evaluasi untuk mengetahui tingkat akurasi model serta faktor yang paling berpengaruh terhadap kelulusan(Bawani et al., 2024).

4. HASIL DAN PEMBAHASAN

Penerapan Data Mining Pada RapidMiner

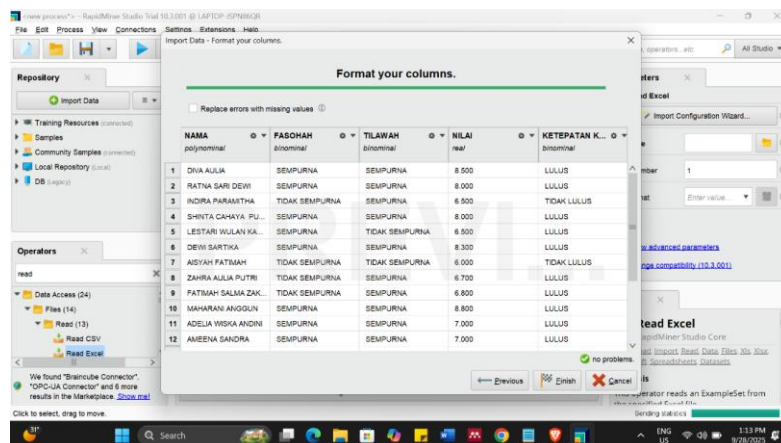
Tahapan ini diawali dengan pemilihan data lalu penentuan tabel kelas seperti dibawah ini:



Gambar 1. Penerapan Data Mining Pada RapidMiner.

Dalam penelitian ini, data yang dimanfaatkan merupakan DATA AL-BARQI.xlsx yang diolah menggunakan RapidMiner Studio. Proses diawali dengan impor data melalui operator Read Excel, kemudian dilakukan preprocessing berupa penanganan data kosong, normalisasi, serta pemisahan Data pelatihan dan data pengujian digunakan bersama dengan algoritma Random Forest dipilih sebagai metode klasifikasi kelulusan santri. Metode ini membangun beberapa decision tree serta menggabungkan hasilnya agar menghasilkan prediksi yang lebih tepat dan konsisten. Hasil implementasi menunjukkan bahwa algoritma Random Forest menunjukkan kinerja yang optimal. Hal ini membuktikan bahwa metode ensemble learning ini efektif digunakan untuk memprediksi kelulusan santri dan dapat menjadi dasar evaluasi serta strategi pembelajaran di pesantren.

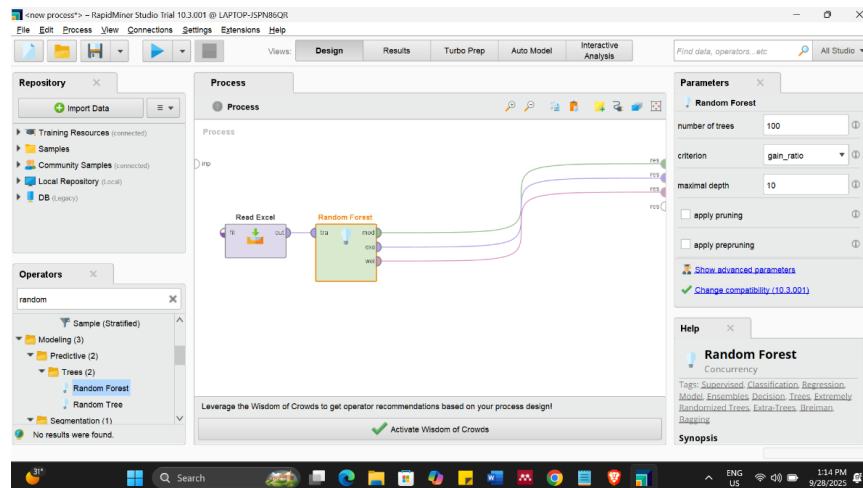
Hasil Pengolahan Data Pada Rapid Miner



Gambar 2. Hasil Pengolahan Data Pada RapidMiner.

Pada tahap awal analisis data, dilakukan proses import data dari file Excel ke dalam platform RapidMiner Studio. Data yang digunakan berisi informasi mengenai nama peserta, tingkat fasahah (kefasihan), tilawah (bacaan Al-Qur'an), nilai numerik, serta ketepatan kriteria kelulusan. Dalam proses import, RapidMiner secara otomatis mendeteksi tipe data masing-masing kolom. Kolom "NAMA" diidentifikasi sebagai polynominal (teks), kolom "FASOHAH", "TILAWAH", dan "KETEPATAN KELULUSAN" diidentifikasi sebagai binominal (kategorikal dengan dua nilai: "SEMPURNA" / "TIDAK SEMPUKURNA" atau "LULUS" / "TIDAK LULUS"), sedangkan kolom "NILAI" diidentifikasi sebagai real (numerik). Hal ini menunjukkan bahwa RapidMiner telah berhasil memahami struktur data secara tepat, sehingga tidak diperlukan penyesuaian manual pada tipe data. Proses import berjalan tanpa kendala, seperti yang ditunjukkan oleh notifikasi "no problems." di bagian bawah jendela. Data yang terimpor mencakup 12 baris (peserta) dengan 5 kolom utama, siap untuk tahap preprocessing dan analisis lebih lanjut. Langkah ini merupakan fondasi penting dalam proses data mining, karena akurasi dan kebersihan data akan sangat memengaruhi hasil analisis selanjutnya. Dengan data yang sudah terstruktur dengan benar, tahap berikutnya seperti transformasi, pemodelan, dan evaluasi — dapat dilakukan dengan lebih efisien dan akurat.

Implementasi

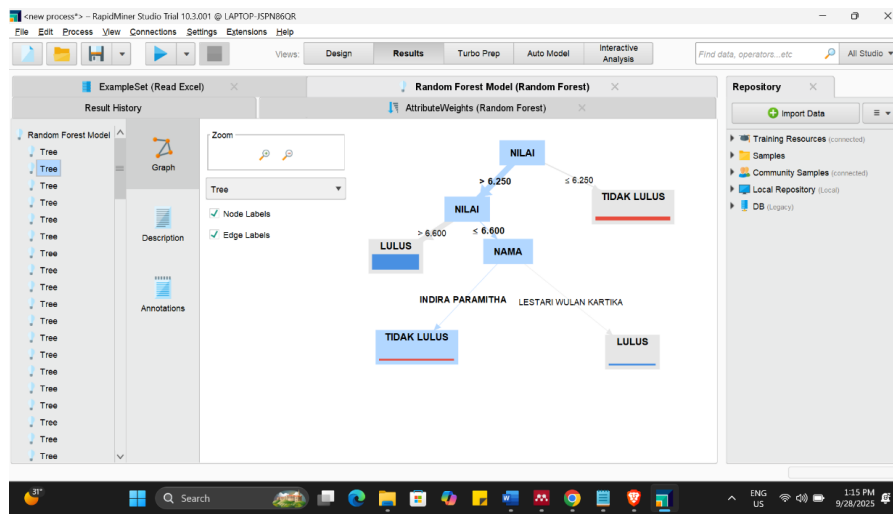


Gambar 3. Implementasi.

Setelah data berhasil diimpor, dilakukan proses pemodelan menggunakan algoritma Random Forest. Pada konfigurasi model, ditetapkan jumlah pohon (number of trees) sebanyak 100, kriteria pemisahan (criterion) menggunakan gain_ratio, dan kedalaman maksimal (maximal depth) sebesar 10. Fitur pruning dan prepruning tidak diaktifkan untuk mempertahankan kompleksitas model. Algoritma Random Forest dipilih karena kemampuannya dalam menghindari overfitting. Proses ini menjadi langkah penting dalam

membangun model prediktif yang akan digunakan untuk menganalisis tingkat berdasarkan parameter fasohah, tilawah, dan nilai.

Hasil Evaluasi Random Forest



Gambar 4. Hasil Visualisasi Model Random Forest.

Setelah proses pemodelan, hasil Random Forest divisualisasikan dalam bentuk pohon keputusan. Dari struktur pohon terlihat bahwa “NILAI” menjadi atribut pembagi utama, dengan ambang batas > 6.250 sebagai penentu awal kelulusan. Jika nilai ≤ 6.250 , peserta langsung diklasifikasikan sebagai TIDAK LULUS. Bagi peserta dengan nilai > 6.250 , proses klasifikasi dilanjutkan dengan mempertimbangkan nilai > 6.600 . Jika memenuhi, maka statusnya LULUS. Namun jika tidak, maka keputusan ditentukan berdasarkan atribut “NAMA”, yang menunjukkan bahwa model juga mempertimbangkan faktor individu dalam pengambilan keputusan — meskipun secara statistik, ini bisa menjadi indikator overfitting atau noise. Visualisasi ini membantu memahami logika pengambilan keputusan model, sekaligus menunjukkan bahwa nilai akademik (NILAI) merupakan faktor dominan dalam menentukan kelulusan.

5. KESIMPULAN DAN SARAN

Kesimpulannya adalah penerapan metode data mining dengan algoritma Random Forest pada program pembelajaran Al-Barqi di Pesantren Salafiyah Syafi'iyah mampu memberikan hasil prediksi kelulusan santri secara akurat dan reliabel. Proses pengolahan data menggunakan RapidMiner melalui tahapan preprocessing, pemodelan, dan implementasi berhasil menghasilkan model klasifikasi yang efektif dalam menentukan status kelulusan santri. Faktor yang paling dominan memengaruhi kelulusan adalah nilai numerik, dengan ambang batas tertentu yang menjadi penentu utama, diikuti oleh aspek fasohah dan tilawah. Hasil ini

membuktikan bahwa metode ensemble learning seperti Random Forest dapat meningkatkan objektivitas dan transparansi dalam sistem evaluasi santri. Selain itu, penelitian ini memberikan kontribusi penting dalam upaya digitalisasi sistem penilaian pesantren sehingga dapat menjadi dasar bagi pengelola dalam merancang startegi pembelajaran yang lebih tepat sasaran dan akuntabel.

DAFTAR REFERENSI

- Abdul Khalim, K., Hayati, U., & Bahtiar, A. (2023). Perbandingan prediksi penyakit hipertensi menggunakan metode Random Forest dan Naïve Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 498–504. <https://doi.org/10.36040/jati.v7i1.6376>
- Bawani, M. A. F., Ashari, M. Y., & Wardani, I. K. (2024). Implementasi pembelajaran Pendidikan Agama Islam berbasis Kurikulum Merdeka di SMPN 1 Peterongan Jombang. *QAZI: Journal of Islamic Studies*, 1(1), 1–11.
- Diantika, S., Nalatissifa, H., Maulidah, N., Supriyadi, R., & Fauzi, A. (2024). Penerapan teknik random oversampling untuk memprediksi ketepatan waktu lulus menggunakan algoritma Random Forest. *Computer Science (CO-SCIENCE)*, 4(1), 11–18. <https://doi.org/10.31294/coscience.v4i1.1996>
- Firdausiyah Luluk, L., Nasa'i, I., & Bakar, M. Y. A. (2022). Sistem evaluasi pembelajaran kurikulum kepesantrenan di Pondok Pesantren Darul 'Ulum Jombang. *Al-Muttaqin: Jurnal Studi, Sosial, dan Ekonomi*, 3(1), 10–19.
- Fitriana, F., Mirani, D., & Dianti, P. (2020). Teaching and learning to read and write the Qur'an using Al-Barqy method. *Conciencia*, 20(1), 14–27. <https://doi.org/10.19109/conciencia.v20i1.4462>
- Magriza, I. A. (2022). *Analisis penerapan algoritma pada preprocessing data*.
- Masnawati, E., & Fitria, S. N. (2024). Peran Taman Pendidikan Al-Qur'an (TPQ) dalam pengembangan akhlak anak. *Irsyaduna: Jurnal Studi Kemahasiswaan*, 4(2), 213–224. <https://doi.org/10.54437/irsyaduna.v4i2.1738>
- Oon Wira Yuda, Darmawan, T., Lim Sheih Yee, & Susanti. (2022). Penerapan data mining untuk klasifikasi kelulusan mahasiswa tepat waktu menggunakan metode Random Forest. *SATIN – Sains dan Teknologi Informasi*, 8(2), 122–131. <https://doi.org/10.33372/stn.v8i2.885>
- Sihabuddin, M., Maarif, I. B., & Roziqin, M. K. (2024). Peran pondok pesantren dalam membentuk akhlak santri. *Jurnal Pendidikan Sosial dan Humaniora*, 3(3), 2280–2294.
- Sobari, S., Purnamasari, A. I., Bahtiar, A., & Kaslani, K. (2025). Meningkatkan model prediksi kelulusan santri tahfidz di Pondok Pesantren Al-Kautsar menggunakan algoritma Random Forest. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5704>
- Vibrianti, V., Wahyudin, E., Kaslani, K., Pratama, D., & Dwilestari, G. (2024). Klasifikasi barang produksi pada TNT Guitar Workshop dengan metode Naive Bayes menggunakan Rapid Miner. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1432–1438. <https://doi.org/10.36040/jati.v8i2.8966>

Warahmah, M., Risnita, & Jailani, M. S. (2023). Pendekatan dan tahapan penelitian dalam kajian pendidikan anak usia dini. *Jurnal DZURRIYAT: Jurnal Pendidikan Islam Anak Usia Dini*, 1(2), 72–81. <https://doi.org/10.61104/jd.v1i2.32>