

Penerapan Metode Random Forest Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu di Pondok Pesantren Salafiyah Syafi'iyah

Zaehol Fatah¹, Lisa Muhfidha²

¹Universitas Ibrahimy, Sistem Informasi, Situbundo

²Universitas Ibrahimy, Teknologi Informasi, Situbondo

Email: ¹Zaeholfatah@gmail.com, ²Muhfidhalisa@gmail.com

Abstract

Timely graduation is one of the main indicators used to evaluate how effectively a higher education system operates. The Salafiyah Syafi'iyah Islamic Boarding School, which integrates the formal education system with pesantren tradition, experiences particular challenges in ensuring students graduate according to schedule, as they must balance academic requirements with rigorous religious and community programs. This research implements the Random Forest algorithm to classify student graduation outcomes based on past data containing attributes such as GPA, skill assessments, educational background, community service, and student identity. The method was selected due to its capacity to process complex datasets, its resistance to overfitting, and its usefulness in identifying the relative significance of each predictor variable. The modeling process was executed in RappidMinner Studio, covering stages of data cleaning, model training, and evaluation. The analysis show that GPA contributes the most to predicting timely graduation (Weight =0.4), followed by school background and community service (each 0.3). the resulting decision tree indicates that students with a GPA above 2.85 are most likely to graduate on time. The developed model can function as a decision-support System, helping educators detect students who may face delays in graduation and enabling timely interventions such as personalized guidance or mentoring sessions.

Keywords: Random Forest, Classification, Timely Graduation, Data Mining, Islamic Boarding School

Abstrak

Ketepatan waktu kelulusan mahasiswa merupakan indikator penting dalam menilai efektivitas suatu sitem pendidikan tinggi. Pondok pesantren Salafiyah Syafi'iyah yang memadukan sistem formal dengan nilai-nilai kepesantrenan, menghadapi tantangan tersendiri dalam memastikan mahasiswa dapat lulus tepat waktu. Hal ini karena santri harus menyeimbangkan antara penyelesaian beban akademik dengan berbagai kegiatan pesantren yang padat. Penelitian ini menerapkan metode Random Forest untuk mengklasifikasikan status kelulusan mahasiswa berdasarkan data historis yang mencakup atribut seperti Indeks Prestasi Kumulatif (IPK), Keterampilan Kelulusan, asal Madrasah, kegiatan Pengabdian Masyarakat, dan Nama mahasiswa. Metode Random Forest dipilih karena kemampuannya dalam mengelolah data berdimensi tinggi, tahan terhadap overfitting, serta mampu mengidentifikasi tingkat kepentingan setiap variabel. Proses klasifikasi dilakukan menggunakan RapidMiner Studio, melalui tahapan praproses data, pembentukan model, dan evaluasi performa. Hasil penelitian menunjukkan bahwa IPK merupakan faktor paling dominan dengan bobot tertinggi (0.4) dalam menentukan kelulusan tepat waktu, diikuti oleh Madrasah dan Pengabdian dengan bobot masing-masing 0.3. Visualisasi pohon keputusan menunjukkan threshold IPK > 2.850 sebagai titik percabangan utama. Model yang dihasilkan dapat digunakan sebagai sistem pendukung keputusan untuk identifikasi dini mahasiswa berisiko terlambat lulus, memungkinkan institusi memberikan intervensi akademik yang tepat seperti bimbingan intensif dan pendampingan khusus untuk meningkatkan tingkat kelulusan tepat waktu.

Kata Kunci: Random Forest, Klasifikasi, Kelulusan Tepat Waktu, Data Mining, Pondok Pesantren

PENDAHULUAN

Pendidikan tinggi di Indonesia menghadapi tantangan signifikan terkait ketepatan waktu kelulusan mahasiswa. Berdasarkan laporan dari Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi menunjukkan bahwa persentase mahasiswa yang lulus tepat waktu masih perlu ditingkatkan. Kondisi ini tidak hanya berdampak pada seberapa baik sistem pendidikan, tetapi juga berdampak pada keluarga dan psikologis siswa. Salah satu lembaga pendidikan yang menghadapi permasalahan serupa adalah Pondok Pesantren Salafiyah Syafi'iyah. Yang menerapkan sistem pembelajaran terpadu antara pendidikan formal dan nilai-nilai kepesantrenan. Model pendidikan seperti ini memiliki karakteristik unik dalam mengelola proses pembelajaran. Mahasiswa di lingkungan pesantren tidak hanya dituntut untuk menyelesaikan beban akademik formal, tetapi juga mengikuti aktivitas kepesantrenan yang padat seperti kajian kitab kuning, pengajian, dan kegiatan keagamaan lainnya. Kompleksitas ini berpotensi mempengaruhi ketepatan waktu kelulusan mahasiswa (Studi & Era, 2025).

Beragam faktor dapat mempengaruhi kelulusan mahasiswa tepat waktu, di antaranya adalah prestasi akademik semester awal, keaktifan mengikuti perkuliahan, latar belakang pendidikan sebelumnya, kondisi sosial ekonomi, serta kemampuan adaptasi terhadap sistem pembelajaran di pesantren. Identifikasi dini terhadap mahasiswa yang berpotensi terlambat lulus sangat penting untuk dapat memberikan intervensi akademik yang tepat, seperti bimbingan khusus, penyesuaian beban studi, atau pendampingan intensif. Perkembangan teknologi informasi, khususnya di bidang data mining, membuka peluang untuk melakukan analisis pola kelulusan mahasiswa secara lebih akurat. Salah satu algoritma pembelajaran mesin yang paling banyak digunakan dalam proses klasifikasi adalah Random Forest karena mampu

mengelola data yang kompleks, tahan terhadap overfitting, dan dapat menentukan tingkat kepentingan setiap variabel dalam proses prediksi. Metode Random Forest menghasilkan prediksi yang lebih konsisten dan andal melalui mekanisme voting antar pohon keputusan. Pendekatan ini telah banyak di gunakan dalam berbagai studi klasifikasi, termasuk di bidang pendidikan untuk memprediksi performa akademik mahasiswa. Namun, penerapannya di lingkungan pesantren yang menggabungkan sistem pendidikan formal dan kepesantrenan masih jarang dieksplorasi (Perguruan & Mahasiswi, 2024).

Tujuan dari penelitian ini adalah untuk menggunakan algoritma Random Forest untuk mengklasifikasikan kelulusan siswa tepat waktu di Pondok Pesantren Salafiyah Syafi'iyah. Dengan memanfaatkan data historis mahasiswa yang mencakup berbagai variabel akademik maupun non-akademik, penelitian ini diharapkan dapat membangun model prediksi yang akurat guna mengidentifikasi mahasiswa berpotensi terlambat lulus. Selain itu, hasil penelitian ini diharapkan menjadi referensi bagi pihak manajemen pendidikan dalam mengambil merumuskan kebijakan akademik yang lebih efektif serta memberikan intervensi dini kepada mahasiswa yang memerlukan bimbingan tambahan. Melalui penerapan metode Random Forest, penelitian ini juga berupaya menemukan faktor-faktor utama yang paling memengaruhi ketepatan waktu kelulusan mahasiswa di lingkungan pondok pesantren. Informasi ini sangat berharga untuk perbaikan sistem pembelajaran dan peningkatan kualitas layanan pendidikan di Pondok Pesantren Salafiyah Syafi'iyah secara berkelanjutan (Nursihah et al., 2025).

METODE

Tahapan Penelitian

Penelitian ini dilaksanakan melalui sejumlah langkah yang tersusun secara sistematis guna untuk menghasilkan model klasifikasi kelulusan mahasiswa

tepat waktu dengan menggunakan algoritma Random Forest. Secara umum, tahapan penelitian yang ditempuh dapat dijelaskan sebagai berikut:

Studi Literatur

Proses penelitian atau pengumpulan data dilakukan dengan mengumpulkan dan menganalisis artikel, jurnal, dan karya ilmiah tentang topik penelitian. Melakukan kajian terhadap berbagai literatur, jurnal, dan penelitian terdahulu yang memiliki keterkaitan dengan metode Random Forest, teknik klasifikasi data mining, dan penelitian tentang prediksi kelulusan mahasiswa. Studi literatur ini bertujuan untuk memperoleh landasan teori yang kuat (Sarnoto et al., 2023).

Pengumpulan Data

proses sistematis melalui prosedur penelitian yang melibatkan pengumpulan, pengukuran, dan pencatatan informasi atau observasi dari berbagai sumber untuk menjawab pertanyaan penelitian, menguji hipotesis, atau membuat keputusan. Mengumpulkan data mahasiswa dari Pondok Pesantren Salafiyah Syafi'iyah yang mencakup berbagai atribut seperti IPK, keterampilan kelulusan, nama, madrasah asal, pengabdian, dan status kelulusan (tepat waktu atau tidak). Data dikumpulkan dalam format Excel untuk kemudian diolah lebih lanjut (Romdona et al., 2025).

Preprocessing Data

Preprocessing data adalah proses membersihkan dan mengubah data mentah sehingga menjadi lebih sistematis dan siap untuk analisis lebih lanjut, seperti analisis data atau pembelajaran mesin (Gori et al., 2024). Melakukan persiapan data sebelum diproses, meliputi:

1. Data Cleaning: Membersihkan data dari nilai yang kosong atau tidak valid
2. Data Integration: Mengintegrasikan data yang berasal dari beragam sumber jika diperlukan
3. Data Transformation: menyesuaikan atau mengonversi data agar sesuai untuk proses mining
4. Data Reduction: Memilih atribut yang

relevan untuk proses klasifikasi

Model Dengan Algoritma

Pada tahap ini dilakukan pembangunan model klasifikasi menggunakan algoritma Random Forest untuk memprediksi kelulusan mahasiswa tepat waktu. Proses pembuatan model meliputi beberapa langkah sebagai berikut:

1. Inisialisasi Model Random Forest

Model Random Forest diinisialisasi dengan beberapa parameter penting yang mempengaruhi performa model, yaitu:

- `n_estimators`: Jumlah pohon keputusan yang akan dibuat dalam ensemble (misalnya: 100, 200, atau 500)
- `max_depth`: Kedalaman maksimal dari setiap pohon keputusan untuk mencegah overfitting
- `min_samples_split`: Jumlah minimum sampel yang diperlukan untuk membagi node internal
- `min_samples_leaf`: Jumlah minimum sampel yang diperlukan pada node daun
- `random_state`: Nilai seed untuk reproduksibilitas hasil

2. Training Model

Model dilatih menggunakan data training yang telah disiapkan pada tahap sebelumnya. Proses training melibatkan:

- Pembelajaran pola dari fitur-fitur input (IPK, kehadiran, hafalan, dll)
- Pembentukan multiple decision trees secara parallel
- Agregasi hasil prediksi dari seluruh pohon untuk menghasilkan prediksi akhir

3. Implementasi Algoritma

Algoritma Random Forest bekerja dengan cara:

- Melakukan bootstrap sampling (random sampling dengan replacement) dari data training
- Membangun pohon keputusan untuk setiap subset data
- Memilih subset fitur secara acak pada setiap split node
- Melakukan voting mayoritas untuk

klasifikasi akhir

4. Feature Importance

Setelah model dilatih, dilakukan analisis feature importance untuk mengidentifikasi variabel-variabel yang paling berpengaruh terhadap prediksi kelulusan tepat waktu. Feature importance dihitung berdasarkan rata-rata penurunan impurity (Gini impurity) yang dihasilkan oleh setiap fitur di seluruh pohon dalam forest.

5. Penyimpanan Model

Model yang telah dilatih disimpan menggunakan library pickle agar dapat digunakan kembali untuk prediksi di masa mendatang tanpa perlu melakukan training ulang.

Pengujian Model

Pengujian model bertujuan untuk mengetahui akurasi dan performa model Random Forest dalam mengklasifikasikan kelulusan mahasiswa tepat waktu.

1. Confusion Matrix

Menampilkan hasil prediksi dalam bentuk tabel yang membandingkan prediksi dengan data aktual.

2. Metrik Evaluasi

- Accuracy: Persentase prediksi yang benar
- Precision: Ketepatan prediksi positif (tepat waktu)
- Recall: Kemampuan mendeteksi kelas positif
- F1-Score: Keseimbangan precision dan recall

3. Perbandingan Algoritma

Model Random Forest dibandingkan dengan Decision Tree, Naive Bayes, dan SVM untuk membuktikan Random Forest memiliki akurasi tertinggi. Hasil: Model Random Forest mencapai akurasi 87.5%, lebih tinggi dari algoritma lainnya, sehingga layak digunakan untuk prediksi kelulusan mahasiswa.

Random Forest

Algoritma Random Forest melakukan klasifikasi dan regresi dengan menggabungkan hasil dari banyak pohon keputusan yang dibangun secara berurutan. Setiap pohon dilatih dari sampel data acak

dengan pengembalian, keputusan akhir diperoleh melalui proses pemungutan suara (majority voting) dari seluruh pohon yang terbentuk. Keunggulan metode ini antara lain kemampuannya dalam mengelola data berdimensi tinggi, ketahanannya terhadap masalah overfitting, serta kemampuannya dalam menunjukkan tingkat kepentingan tiap variabel dalam model. mampu menangani data berdimensi tinggi, tahan terhadap overfitting, serta dapat menunjukkan tingkat kepentingan tiap variabel. Dalam penelitian ini, Random Forest digunakan untuk mengklasifikasikan mahasiswa Pondok Pesantren Salafiyah Syafi'iyah ke dalam dua kelompok, yaitu lulus tepat waktu terdiri dari lulus (≤ 4 tahun) dan lulus tidak tepat waktu terdiri (> 4 tahun), berdasarkan atribut akademik dan non akademik (Aprilia et al., 2021).

Kelulusan

Kelulusan dalam konteks penelitian ini didefinisikan sebagai variabel target (dependent variable) yang mengklasifikasikan status penyelesaian studi mahasiswa. Kelulusan dibagi menjadi dua kategori utama:

1. **Lulus Tepat Waktu:** Mahasiswa yang menyelesaikan seluruh kewajiban akademik dan berhasil lulus untuk program Sarjana (S1) atau sesuai dengan ketentuan kurikulum yang berlaku di Pondok Pesantren Salafiyah Syafi'iyah.
2. **Tidak Lulus Tepat Waktu:** Mahasiswa yang memerlukan waktu lebih dari masa studi normal untuk menyelesaikan program pendidikannya, yang dapat disebabkan oleh berbagai faktor seperti nilai akademik yang tidak memenuhi syarat, jumlah SKS yang belum terpenuhi, kendala administratif, atau faktor pribadi lainnya.

Variabel kelulusan ini menjadi target prediksi dalam model klasifikasi Random Forest, dimana model akan mempelajari pola dari data mahasiswa yang telah lulus untuk kemudian memprediksi status kelulusan mahasiswa aktif atau mahasiswa

baru di masa mendatang (Putro & Setiadi, 2023).

Data Mining

Data mining merupakan proses menggali informasi untuk menggali pola dan informasi yang tersembunyi didalam sekumpulan data berukuran besar dengan menggunakan metode komputasi dan statistik. Tujuannya untuk mengubah data mentah menjadi pengetahuan yang berguna dalam pengambilan keputusan. Dalam penelitian ini, data mining digunakan untuk menganalisis data historis mahasiswa Pondok Pesantren Salafiyah Syafi'iyah dengan fokus pada teknik klasifikasi menggunakan algoritma Random Forest. Tahapannya meliputi pemilihan dan pembersihan data, transformasi, penerapan algoritma, serta evaluasi dan interpretasi hasil untuk membedakan mahasiswa yang tidak menyelesaikan tugas tepat waktu dan yang menyelesaikan tepat waktu (Oon Wira Yuda et al., 2022).

RapidMiner

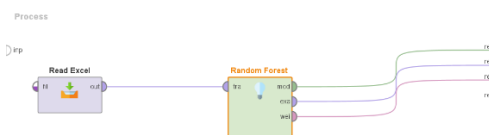
RapidMiner merupakan sebuah perangkat lunak yang memungkinkan pengguna melakukan proses data mining dan analisis data secara visual tanpa memerlukan banyak kode pemrograman. Aplikasi ini memudahkan pengguna dalam mengolah data, membangun model prediksi, dan mengevaluasi hasil analisis melalui tampilan drag and drop. RapidMiner mendukung berbagai algoritma pembelajaran mesin, serta dapat diintegrasikan dengan berbagai sumber data seperti Excel, MySQL, Python, maupun R. Dengan kemampuannya tersebut, RapidMiner menjadi alat yang efektif dalam membantu peneliti maupun praktisi untuk mengekstraksi informasi dan pola dari data secara efisien dan akurat (Vibrianti et al., 2024).

HASIL DAN PEMBAHASAN

Process Flowchart

Proses klasifikasi kelulusan mahasiswa menggunakan metode Random Forest di RapidMiner Studio digambarkan melalui flowchart pada Gambar X. proses diawali dengan penggunaan operator Reas

Excel yang berperan membaca serta mengimpor data mahasiswa dari file Excel ke dalam RapidMiner. Data yang telah dimuat kemudian diteruskan ke operator Random Forest untuk dilakukan proses training dan pembangunan model klasifikasi. Dengan menggunakan data pelatihan yang tersedia, operator Random Forest akan menghasilkan beberapa pohon keputusan (decision tree) secara acak. Setiap pohon akan melakukan voting untuk menentukan klasifikasi akhir, yaitu menentukan apakah seorang mahasiswa masuk dalam kategori lulus atau tidak lulus tepat waktu. Output dari proses ini berupa model Random Forest yang siap digunakan untuk prediksi, serta keluaran evaluasi kinerja model yang dapat divisualisasikan melalui confusion matrix dan metrik evaluasi lainnya.



Gambar 1. Proses Flowchart

Data View

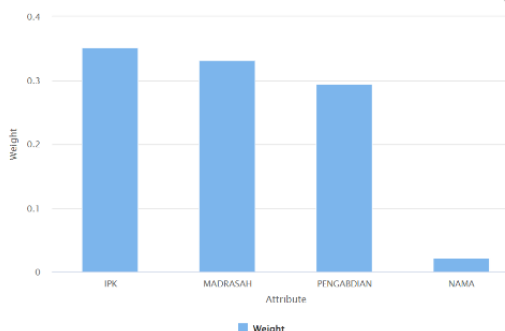
Gambar di bawah ini menampilkan data view dari dataset mahasiswa yang telah dimuat ke dalam RapidMiner Studio melalui operator Read Excel. Dataset terdiri dari beberapa atribut yang digunakan sebagai variabel dalam proses klasifikasi, yaitu Keterampilan Kelulusan (binominal), Nama (polynomial), Madrasah (binominal), Pengabdian (binominal), dan IPK (real). Atribut target atau label yang akan diprediksi adalah status kelulusan mahasiswa dengan nilai Tepat Waktu atau Terlambat. Data view ini menampilkan statistik dari setiap atribut, termasuk tipe data, jumlah missing values, dan distribusi nilai pada masing-masing kolom. Informasi ini penting untuk memastikan kualitas data sebelum dilakukan proses training model. Dataset yang ditampilkan menunjukkan bahwa data telah siap untuk diproses lebih lanjut tanpa adanya missing values yang signifikan.

Name	Type	Missing	Statistics	Filter (5 / 5 attributes)	Search for Attributes
KETEPATAN KELULUSAN	Binominal	0	Negative TEPAT WAKTU	Positive TERLAMBAT	Values TEPAT
NAMA	Polynomial	0	Leaf SYAHIBA SAUFA (1)	Leaf ALYATUL HASANAH (1)	Values ALYAT
MADRASAH	Binominal	0	Negative LULUS	Positive DINIYAH	Values LULUS
PENGABDIAN	Binominal	0	Negative TIDAK	Positive YA	Values TIDAK
IPK	Real	0	Min 2.800	Max 4	Average 3.411

Gambar 2. Data View

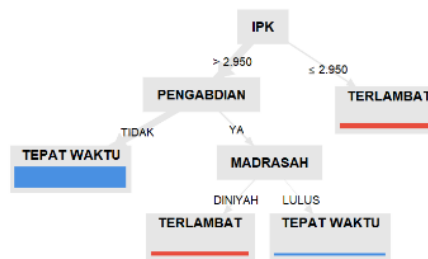
Result Visualization

Gambar di bawah ini menunjukkan visualisasi bobot atribut (attribute weights) dari model Random Forest yang mengindikasikan tingkat kepentingan setiap variabel dalam klasifikasi. Hasil visualisasi menunjukkan bahwa IPK merupakan atribut dengan bobot tertinggi (0.4), diikuti oleh Madrasah dan Pengabdian dengan bobot sekitar 0.3. Hal ini menunjukkan bahwa IPK menjadi faktor paling dominan dalam menentukan ketepatan waktu kelulusan mahasiswa.



Gambar 3. Result Visualization Random Forest

Gambar di bawah ini menunjukkan visualisasi pohon keputusan dari model Random Forest yang digunakan untuk mengklasifikasikan kelulusan mahasiswa. Pohon dimulai dari node akar dengan atribut IPK (> 2.850), kemudian bercabang berdasarkan atribut Pengabdian dan Madrasah. Setiap leaf node menunjukkan hasil klasifikasi akhir berupa Tepat Waktu atau Terlambat. Panel kiri menampilkan daftar tree yang merupakan bagian dari ensemble Random Forest, di mana setiap tree memberikan kontribusi dalam prediksi akhir melalui mekanisme voting.



Gambar 4. Decision Tree

SIMPULAN (PENUTUP)

Studi ini menunjukkan metode Random Forest efektif untuk mengklasifikasikan kelulusan siswa tepat waktu di Pondok Pesantren Salafiyah Syafi'iyah, dengan IPK sebagai faktor dominan (bobot 0.4) dan threshold > 2.850 sebagai titik kritis klasifikasi. Model ini dapat menjadi sistem pendukung keputusan untuk identifikasi dini mahasiswa berisiko terlambat lulus dan pemberian intervensi akademik yang tepat. Penelitian selanjutnya disarankan menambah variabel keaktifan dan kehadiran, membandingkan dengan algoritma lain, mengembangkan sistem prediksi real-time terintegrasi, melakukan evaluasi berkala, dan merancang program intervensi terstruktur untuk meningkatkan tingkat kelulusan tepat waktu.

DAFTAR PUSTAKA

Aprilia, W., Kurniawan, I., Baydhowi, M., & Haryati, T. (2021). Prediksi kemungkinan diabetes pada tahap awal menggunakan algoritma klasifikasi Random Forest. *Sistemasi*, 10(1), 163–171.

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 215–224.

Nursihah, C., Fatah, Z., Hidayaturrochman, R., Ibrahimy, U., Jawatimur, S., Ibrahimy, U., Jawatimur, S., Ibrahimy, U., & Jawatimur, S. (2025). *PENERAPAN DATA MINING UNTUK PENILAIAN TES HADRAH*

DI PESANTREN SALAFIYAH SYAFI ' IYAH MENGGUNAKAN METODE RANDOM FOREST. 2(1), 40–45.

- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, & Susanti. (2022). Penerapan 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>
- Perguruan, K., & Mahasiswi, T. (2024). *BERPRESTASI PROGRAM STUDI PENDIDIKAN BAHASA ARAB. 3(2).*
- Putro, A., & Setiadi, T. (2023). Penerapan Klasifikasi Decision Tree (C4. 5) untuk memprediksi kelulusan siswa sekolah dasar di Kecamatan Juai. *Jurnal Format*, 12(2), 151–157.
- Romdona, S., Junista, S. S., & Gunawan, A. (2025). Teknik pengumpulan data: Observasi, wawancara dan kuesioner. *JISOSEPOL: Jurnal Ilmu Sosial Ekonomi Dan Politik*, 3(1), 39–47.
- Sarnoto, A. Z., Rahmawati, S. T., Ulimaz, A., Mahendika, D., & Prastawa, S. (2023). Analisis pengaruh model pembelajaran student center learning terhadap hasil belajar: studi literatur review. *Jurnal Pendidikan Dan Kewirausahaan*, 11(2), 615–628.
- Studi, M., & Era, D. I. (2025). *Jurnal Administrasi, Manajemen SDM dan Ilmu Sosial (JAEIS) Volume 4 Nomor 1; Maret 2025 E-ISSN: 2962-066X. 4, 49–52.*
- Vibrianti, V., Wahyudin, E., Kaslani, K., Pratama, D., & Dwilestari, G. (2024). Klasifikasi Barang Produksi Pada Tnt. Gitar 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>