



Optimasi Algoritma Logistic Regression untuk Memprediksi Risiko Dropout Mahasiswa dengan Teknik K-Fold

Sandy Hardiansyah¹, Solikhun²

¹Sistem Informasi, STIKOM Tunas Bangsa, Indonesia

²Teknik Informatika, STIKOM Tunas Bangsa, Indonesia

E-Mail: ¹sandy300902hardiansyah@email.com, ²solikhun@amiktunasbangsa.ac.id

Article Info

Article history:

Received Jun 04, 2026

Revised

Accepted

Kata Kunci:

Logistic Regression
Risiko Dropout
Validasi Silang K-Fold
Prediksi Mahasiswa
Evaluasi Klasifikasi

Keywords:

Logistic Regression
Academic Dropout Risk
K-Fold Cross Validation
Predictive Modeling
Performance Evaluation

ABSTRAK

Penelitian ini bertujuan mengembangkan model prediksi untuk mengidentifikasi potensi mahasiswa mengalami putus studi menggunakan algoritma Logistic Regression yang dioptimalkan melalui teknik validasi silang K-Fold. Dataset yang digunakan terdiri atas atribut akademik dan sosial ekonomi mahasiswa yang telah melalui tahap pra-pemrosesan, meliputi penanganan data hilang, transformasi data kategorikal, dan normalisasi fitur. Kinerja model dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa model memiliki performa yang baik dan konsisten, dengan rata-rata akurasi sebesar 83,20% pada K-Fold 5 dan meningkat menjadi 94,60% pada K-Fold 10. Temuan ini menunjukkan bahwa Logistic Regression efektif digunakan sebagai sistem deteksi dini (early warning system) untuk mengidentifikasi risiko putus studi pada mahasiswa. Penelitian ini juga memberikan peluang pengembangan lebih lanjut melalui penambahan fitur, penerapan algoritma pembandingan, serta optimasi model menggunakan teknik ensemble learning dan penyetelan hiperparameter.

ABSTRACT

This research aims to develop a predictive model for identifying students at risk of dropping out by utilizing the Logistic Regression algorithm optimized through the K-Fold Cross-Validation technique. The dataset consists of academic and socioeconomic attributes of students and has undergone preprocessing stages, including missing value handling, categorical data transformation, and feature normalization. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics. The experimental results demonstrate that the proposed model achieves strong and consistent performance, with an average accuracy of 83.20% using 5-Fold Cross-Validation, which increases to 94.60% with 10-Fold Cross-Validation. These findings indicate that Logistic Regression is effective as an early warning system for identifying students at risk of dropping out in higher education institutions. Furthermore, this study provides opportunities for future improvements through the incorporation of additional features, the application of comparative algorithms, and model optimization using ensemble learning techniques and hyperparameter tuning.

This is an open access article under the [CC BY-NC](#) license.



Corresponding Author:

Sandy Hardiansyah,
Sistem Informasi, STIKOM Tunas Bangsa
Jln. Jendral Sudirman Blok A No. 1,2&3 Pematangsiantar, 21127, Sumut, Indonesia
Email: sandy300902hardiansyah@email.com

1. PENDAHULUAN

Perguruan tinggi menjadi pilihan utama bagi masyarakat dalam melanjutkan pendidikan yang mengarah pada penguasaan bidang keahlian tertentu yang diperlukan untuk memasuki dunia kerja. (Masduki et al., 2023). Kondisi ini menegaskan bahwa ketekunan dan konsistensi dalam menjalani pendidikan tinggi sangatlah penting, mengingat tidak semua mahasiswa berhasil menyelesaikan studinya hingga akhir. (Nugroho et al., 2021). Secara umum, putus kuliah diartikan sebagai kondisi di mana seorang mahasiswa yang sebelumnya terdaftar di perguruan tinggi atau universitas menghentikan studi sebelum berhasil menyelesaikan program akademik dan memperoleh gelar yang ditargetkan (Hadi et al., 2025). Di bidang pendidikan, pengambilan keputusan yang didasarkan pada data semakin vital untuk mengidentifikasi kebutuhan siswa, merancang metode pembelajaran, dan meningkatkan mutu hasil Pendidikan (Aprilia et al., 2025). Mahasiswa yang berhenti studi dapat menghambat pertumbuhan ekonomi, menurunkan daya saing, dan mengurangi produktivitas, sehingga berdampak negatif bagi siswa maupun masyarakat secara luas (Ihwal, 2024). Salah satu tantangan yang dihadapi perguruan tinggi adalah adanya mahasiswa yang mengalami putus studi (Yumi et al., 2024). Tingginya biaya pendidikan bagi sebagian masyarakat mendorong pemerintah untuk memberikan bantuan biaya pendidikan kepada yang membutuhkan (Nisa et al., 2024). Untuk mengatasi permasalahan putus studi tersebut, berbagai upaya telah dilakukan, salah satunya dengan memanfaatkan teknologi seperti *machine learning* dalam menganalisis faktor-faktor yang memengaruhi keberlangsungan studi mahasiswa.

Machine learning merupakan istilah dalam ilmu komputer yang merujuk pada cabang yang mempelajari cara merancang algoritma yang dapat belajar atau menyesuaikan diri dengan pola data tanpa harus diprogram secara langsung. *Machine learning* menggunakan berbagai metode komputasi untuk meningkatkan kinerjanya dengan memanfaatkan pengetahuan yang diperoleh dari pengalaman belajar (Kurniawan et al., 2023). *Logistic Regression* adalah metode yang digunakan untuk menganalisis data dengan satu variabel respon (dependen) dan satu atau lebih variabel predictor (Dewi Nasien et al., 2024). *Logistic regression* adalah model statistik yang digunakan untuk mengidentifikasi hubungan antara variabel input dan probabilitas dari hasil output (A'yunan et al., 2023). *K-fold* merupakan salah satu metode yang paling banyak direkomendasikan oleh para peneliti untuk mengatasi kesulitan dalam memperoleh pengklasifikasi yang tepat serta memperkirakan kinerjanya dengan baik, terutama saat menghadapi dataset yang berukuran kecil (Oyedeke, 2023). Model ini dilatih menggunakan $K-1$ lipatan dan diuji pada lipatan yang tersisa. Proses tersebut diulang sebanyak K kali, di mana setiap lipatan berperan sebagai set pengujian satu kali. Tujuan utama dari metode ini adalah untuk mengevaluasi sejauh mana hasil analisis statistik dapat digeneralisasikan ke data independen (Sunarko et al., 2023).

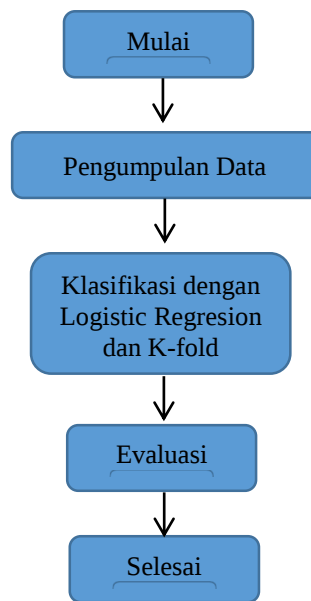
Dalam sebuah penelitian yang telah dilakukan oleh (Bouihi et al., 2024), proses prediksi terhadap kemungkinan mahasiswa mengalami putus kuliah berhasil dilaksanakan dengan menggunakan metode *Logistic Regression*. Penelitian tersebut menunjukkan bahwa penggunaan *Logistic Regression* yang telah diterapkan dengan teknik regularisasi L2 mampu menghasilkan tingkat akurasi pengujian sebesar 81,66%. Selain itu, pendekatan lain yang digunakan dalam penelitian yang sama, yaitu regresi polinomial berganda dengan regularisasi L1-ridge, menghasilkan nilai koefisien determinasi (*R-squared*) sebesar 86,43%, yang menunjukkan kualitas model dalam menjelaskan variabilitas data cukup tinggi. Pada penelitian lain yang telah dilakukan sebelumnya oleh (Kamila & Subastian, 2019), Hasil dari penelitian tersebut menunjukkan bahwa metode *Naïve Bayes* memberikan performa prediksi yang lebih tinggi dengan nilai akurasi sebesar 92%, dibandingkan dengan metode *KNN* yang menghasilkan akurasi prediksi sebesar 85%. Temuan-temuan ini menunjukkan bahwa pemanfaatan algoritma *machine learning* dalam konteks pendidikan, khususnya untuk mendeteksi risiko mahasiswa putus studi, memiliki potensi yang signifikan dalam membantu pengambilan keputusan dan intervensi dini oleh pihak kampus.

Tujuan dari penelitian ini adalah untuk mengoptimalkan kekuatan regularisasi (regularization strength) pada metode *Logistic Regression* guna mendeteksi risiko mahasiswa putus kuliah secara lebih akurat dan efisien, dengan menggunakan pendekatan evaluasi berbasis metode validasi silang *K-Fold* untuk memperoleh model prediksi yang andal dan dapat digeneralisasikan dengan baik.

2. METODE PENELITIAN

2.1 Tahapan Penelitian

Tahapan-tahapan berikut merupakan prosedur yang ditempuh oleh peneliti, yang secara lebih rinci dapat diilustrasikan pada Gambar 1. Diagram alur tersebut merepresentasikan rangkaian langkah sistematis yang dilaksanakan selama proses penelitian, dimulai dari tahap perancangan hingga penyusunan hasil akhir.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Dataset ini diperoleh dari platform open source *Kaggle* (www.kaggle.com) data ini mencakup informasi terkait mahasiswa, seperti identitas, jenis kelamin, Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, status kelulusan (lulus, *dropout*, atau masih aktif), kehadiran, lama studi, dan informasi relevan lainnya. Metode pengumpulan data dilakukan melalui teknik dokumentasi, yaitu dengan memanfaatkan data historis yang telah tersedia tanpa melibatkan proses wawancara atau kuesioner. Data ini dipilih karena relevansi dan kelengkapannya dalam mendukung analisis prediktif untuk risiko *dropout* mahasiswa. Untuk memastikan validitas dan kualitas data, dilakukan seleksi berdasarkan kriteria berikut: hanya data mahasiswa yang status studinya sudah diketahui (lulus atau *dropout*) dan memiliki informasi lengkap pada semua variabel yang diteliti. Data yang tidak memenuhi kriteria ini akan dikeluarkan dari analisis. Sebelum analisis dilakukan, data akan melalui tahap pra-pemrosesan yang meliputi penanganan data yang hilang (dengan pengisian atau penghapusan), transformasi variabel kategorikal menjadi numerik menggunakan teknik seperti label *encoding* atau *one-hot encoding*, normalisasi data numerik (jika diperlukan), dan penentuan variabel target (dalam hal ini status *dropout*, dengan nilai 1 untuk *dropout* dan 0 untuk tidak *dropout*).

2.3 Klasifikasi dengan Logistic Regression dan K-fold

Pada tahap ini, dilakukan proses klasifikasi guna memprediksi potensi mahasiswa mengalami putus studi dengan memanfaatkan algoritma *Logistic Regression*. Pemilihan metode ini didasarkan pada kemampuannya dalam merepresentasikan hubungan antara variabel-variabel independen (seperti Indeks Prestasi Kumulatif, jumlah SKS yang telah ditempuh, dan durasi studi) terhadap probabilitas terjadinya peristiwa biner, yaitu putus studi (*dropout*) atau tidak. Untuk memperoleh hasil evaluasi model yang lebih objektif dan mengurangi kemungkinan bias akibat distribusi data yang tidak merata, digunakan teknik validasi silang *K-Fold*. Dalam penerapannya, data dibagi ke dalam sejumlah lipatan (*folds*) sebanyak K misalnya $K = 5$. Proses pelatihan dilakukan secara berulang sebanyak K kali, di mana pada setiap iterasi, $K - 1$ lipatan digunakan sebagai data latih dan satu lipatan sisanya digunakan sebagai data uji. Dengan demikian, seluruh data berperan secara bergantian sebagai data uji, yang memungkinkan model untuk dievaluasi secara menyeluruh dan menyajikan performa yang lebih representatif terhadap keseluruhan dataset.

Selama proses validasi berlangsung, model *Logistic Regression* dilatih menggunakan parameter regularisasi yang telah ditentukan sebelumnya untuk mencegah *overfitting*. Pada setiap putaran validasi, dihitung sejumlah metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Seluruh nilai metrik dari tiap lipatan kemudian dirata-ratakan untuk memperoleh gambaran performa model secara keseluruhan. Hasil dari tahapan klasifikasi ini menjadi dasar untuk menilai sejauh mana integrasi antara algoritma *Logistic Regression* dan teknik validasi silang *K-Fold* mampu membentuk model prediktif yang akurat, robust, dan dapat digeneralisasikan, khususnya dalam konteks deteksi dini risiko mahasiswa mengalami putus studi.

2.4 Evaluasi

Evaluasi model dilakukan dengan menghitung metrik akurasi, presisi, recall, F1-score, dan *AUC-ROC* pada setiap iterasi *K-Fold Cross-Validation*, yang kemudian dirata-ratakan untuk menilai secara menyeluruh performa algoritma *Logistic Regression* dalam memprediksi risiko mahasiswa putus studi secara akurat dan dapat digeneralisasikan.

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Setelah proses klasifikasi menggunakan algoritma *Logistic Regression* dan validasi silang *K-Fold* dengan nilai $K=5$, diperoleh hasil evaluasi performa model berdasarkan metrik akurasi, presisi, recall, dan *F1-score* pada masing-masing *fold*.

Tabel 1. Hasil Evaluasi Logistic Regression dengan K-Fold (K = 5)

Fold ke-	Akurasi (%)	Presisi (%)	Recall (%)	F1-Score (%)
1	82.00	80.65	78.12	79.37
2	84.50	83.33	81.25	82.28
3	81.00	78.00	80.00	78.99
4	83.50	82.00	84.61	83.29
5	85.00	84.21	83.33	83.77
Rata-rata	83.20	81.64	81.86	81.90

Berdasarkan Tabel 1, dapat dilihat bahwa model *Logistic Regression* menunjukkan performa yang cukup baik dan stabil pada setiap *fold*. Nilai akurasi berkisar antara 81,00% hingga 85,00%, dengan rata-rata sebesar 83,20%. Hal ini mengindikasikan bahwa model mampu mengklasifikasikan sebagian besar data dengan benar. Nilai presisi dan *recall* juga menunjukkan hasil yang seimbang, masing-masing dengan rata-rata sebesar 81,64% dan 81,86%. Artinya, model tidak hanya akurat dalam mengidentifikasi mahasiswa yang benar-benar *dropout* (presisi), tetapi juga mampu menangkap sebagian besar dari mereka yang sebenarnya berisiko *dropout* (*recall*). *F1-score* sebagai gabungan antara presisi dan *recall* menunjukkan nilai rata-rata sebesar 81,94%, yang mendukung kesimpulan bahwa model cukup optimal dalam menangani klasifikasi biner pada kasus ini. Kestabilan hasil antar *fold* yang tidak memiliki deviasi signifikan mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data baru. Hal ini memperkuat validitas penggunaan algoritma *Logistic Regression* yang dikombinasikan dengan teknik validasi silang *K-Fold* dalam mendeteksi risiko mahasiswa putus studi.

3.2 Dataset Menggunakan Kode Python

Kode *Python* ini dimanfaatkan untuk membangun model klasifikasi menggunakan algoritma *Logistic Regression* guna memprediksi potensi mahasiswa mengalami putus studi berdasarkan dataset yang memuat atribut akademik dan sosial ekonomi. Proses diawali dengan tahapan *data preprocessing* yang meliputi pembacaan data, penghapusan atribut identitas yang tidak relevan, transformasi variabel kategorikal menjadi numerik menggunakan teknik *Label Encoding*, serta penanganan nilai hilang (*missing values*) dengan imputasi menggunakan nilai rata-rata. Selanjutnya, seluruh fitur dinormalisasi menggunakan *StandardScaler* untuk memastikan setiap fitur memiliki skala yang seragam demi meningkatkan kinerja model. Evaluasi model dilakukan menggunakan teknik *Stratified K-Fold Cross-Validation* sebanyak 10 lipatan (*fold*) untuk menghindari bias dan *overfitting*, serta memperoleh hasil evaluasi yang representatif. Hasil akurasi dari masing-masing *fold* divisualisasikan dalam bentuk diagram batang yang disertai dengan garis rata-rata akurasi, sehingga memberikan gambaran menyeluruh mengenai performa klasifikasi model terhadap dataset yang digunakan.

```

# Langkah 2: Pra-pemrosesan data
# Drop kolom yang tidak relevan atau terlalu spesifik
data = data.drop(columns=['nim', 'nama'])

# Ubah data kategori menjadi numerik
for column in data.select_dtypes(include='object').columns:
    data[column] = LabelEncoder().fit_transform(data[column])

# Tangani missing value (jika ada)
data = data.fillna(data.mean(numeric_only=True))

# Pisahkan fitur dan target
X = data.drop(columns=['status_putus_studi']) # Fitur
y = data['status_putus_studi'] # Target (0 atau 1)

# Normalisasi fitur
scaler = StandardScaler()
X = scaler.fit_transform(X)

# Langkah 3: Inisialisasi model
model = LogisticRegression(max_iter=1000)

# Langkah 4: Validasi menggunakan K-Fold (k = 10)
kfold = StratifiedKFold(n_splits=10, shuffle=True, random_state=42)
scores = cross_val_score(model, X, y, cv=kfold, scoring='accuracy')

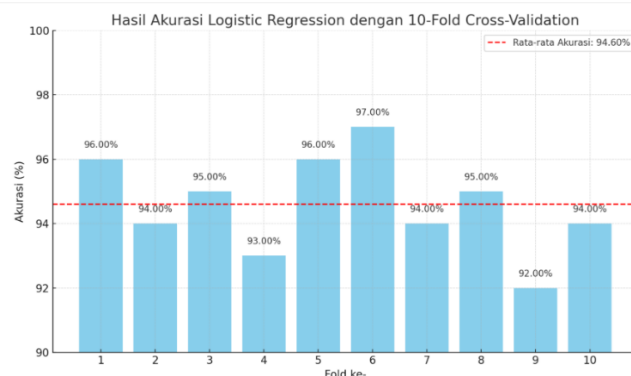
# Cetak hasil validasi
print("Akurasi per fold:", scores)
print("Rata-rata akurasi:", round(scores.mean() * 100, 2), "%")

```

Gambar 2. Menggunakan Kode Python

Gambar 2 memperlihatkan cuplikan kode *Python* yang merepresentasikan tahapan praproses data hingga evaluasi model klasifikasi berbasis algoritma *Logistic Regression*. Tahapan tersebut mencakup penghapusan atribut identitas yang tidak berkontribusi terhadap proses prediksi, transformasi variabel kategorikal menjadi bentuk numerik menggunakan metode *Label Encoding*, penanganan nilai hilang melalui teknik imputasi berbasis rata-rata, serta normalisasi fitur dengan memanfaatkan *StandardScaler* untuk memastikan keseragaman skala data. Selanjutnya, model *Logistic Regression* diinisialisasi dan dievaluasi menggunakan pendekatan *Stratified K-Fold Cross-Validation* sebanyak 10 lipatan guna memperoleh estimasi akurasi rata-rata yang merefleksikan tingkat kemampuan generalisasi model terhadap keseluruhan data yang dianalisis.

3.3 Hasil Diagram Kodingan Menggunakan Kode Python



Gambar 3. Hasil Diagram Menggunakan Kode Python

Diagram tersebut menyajikan visualisasi hasil akurasi dari model *Logistic Regression* yang diimplementasikan pada dataset mahasiswa dengan pendekatan *10-Fold Cross-Validation*, di mana setiap batang mencerminkan tingkat akurasi klasifikasi pada masing-masing *fold* dengan rentang akurasi antara 92% hingga 97%. Garis merah putus-putus yang ditampilkan pada diagram menunjukkan nilai rata-rata akurasi sebesar 94,6%, yang mengindikasikan bahwa data yang digunakan yang terdiri atas atribut akademik dan kondisi sosial ekonomi mahasiswa memiliki struktur dan pola yang cukup representatif dan konsisten dalam mendukung proses klasifikasi mahasiswa berpotensi putus studi secara efektif melalui pendekatan *Logistic Regression*.

4. KESIMPULAN

Hasil penelitian ini mengindikasikan bahwa penerapan algoritma *Logistic Regression* yang dikombinasikan dengan metode validasi silang *K-Fold* mampu memberikan performa prediktif yang andal

dalam mengidentifikasi risiko mahasiswa mengalami putus studi. Evaluasi terhadap parameter kinerja seperti akurasi, presisi, recall, dan *F1-score* menunjukkan bahwa model memiliki tingkat stabilitas dan efektivitas yang tinggi, dengan rata-rata akurasi sebesar 83,20% pada validasi K=5 dan meningkat hingga 94,6% pada K=10. Selain itu, model menunjukkan keseimbangan yang baik antara kemampuan mendeteksi mahasiswa yang benar-benar berisiko (*recall*) dan ketepatan dalam klasifikasi (*precision*), sebagaimana ditunjukkan oleh nilai *F1-score* yang konsisten. Konsistensi performa antar lipatan validasi menunjukkan bahwa model memiliki daya generalisasi yang kuat serta risiko overfitting yang rendah. Dengan demikian, pendekatan ini dapat direkomendasikan sebagai solusi *early warning system* dalam konteks manajemen akademik di perguruan tinggi. Lebih lanjut, temuan ini memberikan landasan untuk pengembangan riset lanjutan, antara lain dengan menambahkan atribut prediktif baru, mengeksplorasi algoritma pembandingan, maupun menerapkan teknik *ensemble learning* dan *hyperparameter tuning* guna meningkatkan kinerja model secara keseluruhan.

REFERENCES

- A'yunan, Y. A. D. K., Indahyanti, U., & Busono, S. (2023). Implementasi Data Mining dalam Klasifikasi Diagnosa Kanker Payudara menggunakan Algoritma Logistic Regression. *Jurnal TEKINKOM*, 6(2), 400–407. <https://doi.org/10.37600/tekinkom.v6i2.948>
- Aprilia, F., Anggraini, R. A., & Putri, Y. D. (2025). *Prediksi Kelulusan Siswa dengan Algoritma Pembelajaran Mesin : Aplikasi Regresi Linear dan Logistik pada Faktor- Faktor Pendidikan Student Graduation Prediction Using Machine Learning Algorithms : Application of Linear and Logistic Regression on Education*. 3(1), 55–64. <https://doi.org/10.25181/rt.v3i1.3897>
- Bouihi, B., Bouselham, A., Aoula, E., Ennibras, F., & Deraoui, A. (2024). Prediction of Higher Education Student Dropout based on Regularized Regression Models. *Engineering, Technology and Applied Science Research*, 14(6), 17811–17815. <https://doi.org/10.48084/etasr.8644>
- Dewi Nasien, Darwin, R., Cia, A., Leo Winata, A., Go, J., M.C, R., Charles Wijaya, R., & Charles Lo, K. (2024). Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression untuk Mengklasifikasi Penyakit Diabetes. *JEKIN - Jurnal Teknik Informatika*, 4(1), 10–17. <https://doi.org/10.58794/jekin.v4i1.640>
- Hadi, I., Ghozali, A., & Handoko, A. R. (2025). *Analisis Efektifitas Feature Selection dalam Pengkayaan Machine Learning untuk Deteksi Dini*. 11(1), 149–157. <https://doi.org/10.26418/jp.v11i1.90362>
- Ihwal, M. (2024). *Klasifikasi Mahasiswa Berpotensi Putus studi menggunakan Algoritma Naive Bayes pada Fakultas Teknik Unismuh Makassar INFO*. 2(1).
- Kamila, V. Z., & Subastian, E. (2019). KNN vs Naive Bayes Untuk Deteksi Dini Putus Kuliah Pada Profil Akademik Mahasiswa. *Jurnal Rekayasa Teknologi Informasi (JURTI)*, 3(2), 116. <https://doi.org/10.30872/jurti.v3i2.3097>
- Kurniawan, R., Wintoro, P. B., Mulyani, Y., & Komarudin, M. (2023). Implementasi Arsitektur Xception Pada Model Machine Learning Klasifikasi Sampah Anorganik. *Jurnal Informatika Dan Teknik Elektro Terapan*, 11(2), 233–236. <https://doi.org/10.23960/jitet.v11i2.3034>
- Masduki, Prihartini, E., & Abdullah, D. (2023). Analisis Faktor-Faktor Yang Mempengaruhi Keputusan Kuliah Di Perguruan Tinggi Swasta. *Jurnal Bisnis Manajemen Dan Kewirausahaan*, 4(1), 205–219. <https://doi.org/10.3917/rpve.441.0093>
- Nisa, I., Bakti, A. M. F., Hermansah, T., & Fanshoby, M. (2024). Peran STF UIN Jakarta Terhadap Tindakan Preventif Putus Kuliah dalam Perspektif Komunikasai Pemabangunan. *Jurnal Ekonomi Bisnis Indonesia*, 19(1), 10–20. <https://www.jurnal.stiebi.ac.id/index.php/Jebi/article/view/447>
- Nugroho, O. F., Permanasari, A., Firman, H., & Riandi, R. (2021). Persepsi Dan Praktik Pendidikan Keberlanjutan Di Indonesia Untuk Education for Sustainability Development (Esd) Selama Pandemi Covid-19. *Eduscience : Jurnal Ilmu Pendidikan*, 7(01). <https://doi.org/10.47007/edu.v7i01.4642>
- Oyedele, O. (2023). Determining the optimal number of folds to use in a K-fold cross-validation: A neural network classification experiment. *Research in Mathematics*, 10(1). <https://doi.org/10.1080/27684830.2023.2201015>
- Sunarko, V. I., Lizard, D., Dimara, S., Sandya, P., Siagian, E., Manalu, D., & Anggraeny, F. T. (2023). *Implementasi K-Fold Dalam Prediksi Hasil Produksi Agrikultur Pada Algoritma K-Nearest Neighbor (KNN)*. 10–16. <https://ejournal.itats.ac.id/integer/article/view/6892>
- Yumi, Lukman, & Bakti, R. (2024). Klasifikasi Mahasiswa Berpotensi Putus Studi Menggunakan Algoritma Decision Tree pada Fakultas Teknik Unismuh Makassar. *Arus Jurnal Sains Dan Teknologi (AJST)*, 2(2).