

Prediksi Kinerja Siswa Pada E-Learning Moodle Platform Menggunakan Algoritma Adaptive Boosting

Jordy Lasmana Putra^{1*}, Agus Subekti²

¹Program Studi Informatika, Fakultas Teknologi Informasi, Universitas Nusa Mandiri
Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar, Jakarta Timur, Indonesia

²Program Studi Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Nusa Mandiri
Jl. Jatiwaringin No. 2, Cipinang Melayu, Makasar, Jakarta Timur, Indonesia

²Badan Riset & Inovasi Nasional (BRIN)
Jl. Sangkuriang 21/154D, Bandung, Indonesia

e-mail: ^{1*}jordy.jlp@nusamandiri.ac.id, ²agus106@brin.go.id, ²agus@nusamandiri.ac.id

Informasi Artikel	Diterima: 27-02-2023	Direvisi: 10-03-2023	Disetujui: 14-03-2023
-------------------	----------------------	----------------------	-----------------------

Abstrak

Pandemi Covid-19 yang sudah berlangsung sejak awal tahun 2020 memberikan dampak besar di berbagai sektor, salah satunya di sektor pendidikan, dimana awalnya pendidikan dilakukan secara tatap muka, karena pandemi mengharuskan proses belajar mengajar dilakukan secara dalam jaringan (daring) Teknologi informasi berkembang sangat pesat dan mempengaruhi berbagai bidang, salah satunya bidang pendidikan, yang dimana pembelajaran secara daring sudah menjadi hal yang biasa untuk era saat sekarang ini, salah satu Learning Management System atau yang sering disingkat LMS yang sering digunakan adalah E-Learning menggunakan platform moodle, ditambah untuk saat pandemic covid-19 proses pembelajaran diarahkan ke sistem daring, sehingga penggunaan E-Learning menjadi meningkat. Melihat hal tersebut penulis bermaksud untuk melakukan penelitian untuk melakukan prediksi terhadap kinerja siswa dalam mengikuti perkuliahan e-learning yang menggunakan moodle platform, penelitian ini melihat dari sisi log activity siswa di moodle platform lalu log tersebut di transformasi agar dapat dilakukan proses klasifikasi oleh algoritma machine learning. Pada penelitian ini penulis melakukan klasifikasi menggunakan algoritma Adaptive Boosting dengan Base Learner C4.5 dengan teknik pra pemrosesan data Resample untuk Imbalance data. Hasil dari penelitian ini didapatkan hasil performansi yang baik, dengan nilai Akurasi 95%, ROC 0.97, dan Kappa 0.90. sehingga penelitian ini dapat menjadi model untuk memprediksi kinerja siswa dengan melihat log aktivitasnya menggunakan platform moodle.

Kata Kunci: Kinerja Siswa, E-Learning, Adaptive Boosting

Abstract

The Covid-19 pandemic, which has been going on since the beginning of 2020, has had a major impact in various sectors, one of which is in the education sector, where initially education was carried out face-to-face, because the pandemic requires the teaching and learning process to be carried out online. Information technology is developing very rapidly and affecting various fields, one of which is the field of education, where online learning has become commonplace for today's era, one of the Learning Management Systems or often abbreviated as LMS that is often used is E-Learning using the moodle platform, plus during the Covid-19 pandemic the learning process is directed to an online system, so that the use of E-Learning becomes increasing. Seeing this, the author intends to conduct research to predict student performance in participating in e-learning lectures using the moodle platform, this study looks at the student activity log on the moodle platform and then the log is transformed so that the classification process can be carried out by machine learning algorithms. In this study, the authors classified using the Adaptive Boosting algorithm with Base Learner C4.5 with the Resample data preprocessing technique for data imbalance. The results of this study obtained good performance results, with an Accuracy value of 95%, ROC 0.97, and Kappa 0.90. So this study can be a model to predict student performance by looking at their activity logs using the Moodle platform.

Keywords: Students Performance, E-Learning, Adaptive Boosting



1. Pendahuluan

Kemajuan teknologi dan Informasi sangat mempengaruhi berbagai bidang, salah satunya adalah bidang pendidikan, hal ini dapat dilihat dari banyaknya penelitian yang dilakukan, seperti penelitian yang dilakukan oleh (Nespereira et al., 2016) yang melakukan perbandingan kinerja algoritma Random Forest dan SVM untuk memprediksi lulus atau tidaknya siswa melihat dari interaksi masa lalu di Learning Management System (LMS).

Penelitian selanjutnya dilakukan oleh (Masangu et al., 2021) dengan melakukan pendekatan machine learning untuk memprediksi prestasi akademik mahasiswa. Pada penelitian yang dilakukan oleh (Irfan et al., 2020) menunjukkan bahwa selama pandemi covid-19 proses pembelajaran dilakukan secara online melalui Learning management System (LMS).

Salah satu Learning Management System (LMS) yang sering digunakan dalam sistem pembelajaran E-Learning ialah menggunakan Moodle Platform (Holiver et al., 2020), pada penelitian yang dilakukan oleh (Chango et al., 2019) melakukan prediksi kinerja akademik mahasiswa dari multi-sumber data dalam lingkungan multimodal dan blended learning menggunakan data fusion dan data mining.

Melakukan prediksi kinerja siswa sangat penting dilakukan, sebagai upaya lembaga pendidikan tinggi untuk meningkatkan kualitas peserta didik (Annisa & Sasongko, 2020), hasil prediksi yang baik dapat didapatkan dengan teknik data mining (Putri et al., 2020), ada banyak metode data mining yang dapat digunakan untuk melakukan sebuah prediksi (Shedriko & Firdaus, 2022) salah satunya adalah algoritma Adaptive Boosting yang merupakan algoritma ensemble learning yang secara konsisten memberikan hasil kinerja klasifikasi yang lebih baik (Sudaryanto & Sudaryanto, 2022)

Untuk itu penulis tertarik untuk melakukan penelitian dengan topic pembahasan yang serupa dari penelitian-penelitian sebelumnya, yaitu mengenai sistem pembelajaran E-Learning, dan mengacu pada penelitian yang dilakukan oleh (Aleksandrova, 2019) yang dimana pada penelitian ini peneliti melakukan prediksi performa mahasiswa di platform moodle dengan melihat log activitynya menggunakan machine learning. Algoritma yang digunakan yaitu: logistic regression, random forest, xgboost, dan neural network, dengan jumlah data 112. menghasilkan akurasi 90.62% dengan algoritma Neural Network dan XgBoost.

Dalam penelitian ini penulis ingin melakukan pengembangan terhadap model

prediksi kinerja siswa berdasarkan log activity mahasiswa dalam menggunakan E-Learning Moodle Platform dengan menggunakan algoritma Adaptive Boosting.

2. Metode Penelitian

2.1. Data Penelitian

Penelitian ini menggunakan dataset Primer dari hasil penelitian penulis di bagian E-Learning dan Biro Administrasi Akademik dan Kemahasiswaan (BAAK) Universitas Bina Sarana Informatika. Tahap preprocessing dilakukan dengan melakukan analisis terhadap data dengan memperhatikan jumlah atribut, jumlah modul dan jumlah cacat pada setiap dataset. Spesifikasi dataset yang digunakan dalam penelitian ini sebagai berikut.

Tabel 1. Spesifikasi *dataset Log Activity E-Learning Students*

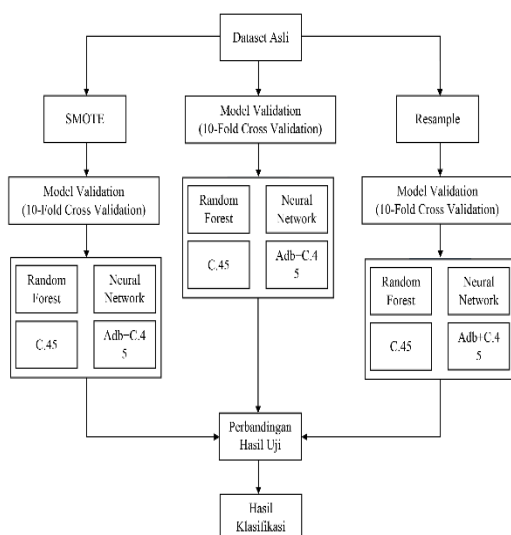
Dataset	Jumlah Atribut	Jumlah Data
Dataset Log Activity E-Learning Students	11	181

Tabel 2. Penjelasan Atribut pada *dataset Log Activity E-Learning Students*

Atribut	Description	Type data
Gender	M=Male, F=Female	Nominal
Enrolment Days After Start Of Semester	[Enrolment Date] – [Date of First Session]	Numeric
First Login	Student First Login in Moodle	Numeric
Last Login	Student Last Login in Moodle	Numeric
Days From First To Last Session	[Last Login] – [First Login]	Numeric
Action Taken	Action Count	Numeric
Avg_Actions per Day	[Action Count] / Avg([Action for the course])	Numeric
Quiz Viewed	Log Quiz Viewed Student in Moodle	Numeric
Forum Viewed	Log Forum Viewed Student in Moodle	Numeric
Files Viewed	Log Files Viewed Student in Moodle	Numeric
Chatting Viewed	Log Chatting Student in Moodle	Numeric
Exam Result	2 levels - 0 (not graduated) and 1 (graduated)	Nominal

2.2. Tahapan Penelitian

Penelitian ini dilakukan dalam beberapa tahapan mulai dari pra-proses data, klasifikasi, validasi, akurasi. Berikut kerangka pemikiran pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan gambar langkah-langkah pemikiran di atas, dapat dijelaskan sebagai berikut:

1. Dataset terlebih dahulu diuji dengan menggunakan algoritma Random Forest, Neural Network, C.45, dan AdaBoost + C4.5 menggunakan 10-fold cross validation.
2. Dataset diuji kembali menggunakan teknik preprocessing data, yaitu SMOTE dan Resample menggunakan Random Forest, Neural Network, C.45, dan AdaBoost + C4.5 menggunakan 10-fold cross validation.
3. Setelah ditemukan hasil terbaik dari hasil uji coba di atas, dilakukan perbandingan terhadap hasil performansi yang didapatkan.
4. Setelah melakukan perbandingan algoritma maka didapat hasil pengujian dengan nilai performa terbaik.

2.2. Resample

Sampling merupakan teknik yang dilakukan dalam mereduksi himpunan data yang sangat besar teknik resampling adalah teknik preprocessing dimana distribusi data diseimbangkan kembali untuk mengurangi efek distribusi kelas yang tidak seimbang dalam proses pembelajaran (Alberth & Soepranoto, 2022).

Teknik resampling disini digunakan untuk memecahkan masalah berupa data yang tidak seimbang. Teknik resampling dilakukan dengan menyeimbangkan data asli berdasarkan serangkaian algoritma sampling dengan menyesuaikan jumlah sampel dalam kelas yang berbeda, kemudian melatih data yang telah seimbang dengan mengangkat algoritma klasifikasi.

Pendekatan resampling dapat dikategorikan menjadi 3 yaitu: metode oversampling, undersampling, dan hibrida yang

merupakan penggabungan dari kedua pendekatan sampling (Lin et al., 2022). Oversampling berfungsi untuk meningkatkan sampel kelas minoritas hingga sama dengan kelas mayoritas lain dengan cara duplikasi secara acak dari sampel kelas minoritas (He et al., 2018)



Sumber: (He et al., 2018)

Gambar 2. Flowchart Oversampling

Sedangkan pada Undersampling memilih secara acak sampel di kelas mayoritas dan menambahkannya ke kelas minoritas dengan membentuk dataset training baru (He et al., 2018)



Sumber: (He et al., 2018)

Gambar 3. Flowchart Undersampling

2.3. AdaBoost

AdaBoost merupakan ensemble learning yang sering digunakan pada algoritma boosting

(Raihan et al., 2022) agar dapat model akurasi tinggi Adaboost bekerja dengan menggabungkan beberapa model akurasi rendah (Bahad & Saxena, 2020). AdaBoost bisa dikombinasikan dengan classifier algoritma lainnya untuk meningkatkan performa (Pebrianti et al., 2022). Secara intuitif, penggabungan beberapa model akan membantu jika model berbeda satu sama lain. Adaboost sukses diterapkan pada beberapa bidang karena dasar teorinya yang kuat, prediksi yang akurat, dan kesederhanaan yang besar (Laila Qadrini et al., 2021)

2.4. C4.5

Algoritma C.45 adalah algoritma klasifikasi data menggunakan teknik decision tree yang terkenal dan disukai karena memiliki banyak kelebihan diantaranya mampu mengolah data numerik (kontinyu) dan diskrit, algoritma ini juga dapat menangani nilai atribut yang hilang, menghasilkan aturan-aturan yang mudah diinterpretasikan dan memiliki kecepatan komputasi yang tinggi diantara algoritma lainnya yang menggunakan memori utama di komputer (Ina et al., 2019)

Algoritma C.45 dipergunakan dalam membentuk decision tree atau pengambilan keputusan. Sebagai salah satu algoritma dalam induksi decision tree yaitu ID3 (Iterative Dichotomiser 3), algoritma C.45 ini memiliki ide dasar yaitu pembuatan pohon keputusan berdasarkan pemilihan atribut yang mempunyai prioritas tertinggi atau memiliki nilai gain tertinggi berdasarkan nilai entropy atribut tersebut sebagai poros atribut klasifikasi. Kemudian secara rekursif cabang-cabang pohon diperluas hingga seluruh pohon terbentuk (Kurniawan, 2018).

3. Hasil dan Pembahasan

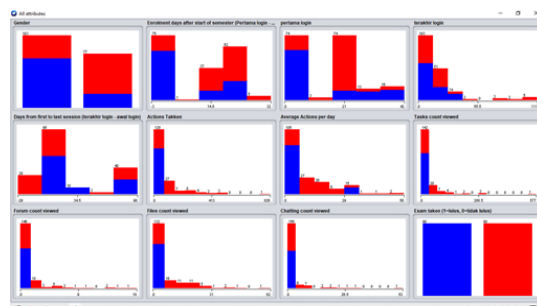
3.1. Preprocessing Data

Tahap selanjutnya dilakukan pengujian dataset dengan algoritma klasifikasi C.45, Random Forest, Neural Network, dan Adaptive Boosting + C4.5. hasil eksperimennya akan dievaluasi menggunakan Accuracy ,confusion matrix dan ROC (Receiver Operating Characteristic) untuk dicari hasil terbaik. Pada penelitian ini tools yang digunakan untuk pengujian algoritma adalah aplikasi WEKA (Waikato Environment for Knowledge Analysis), yang merupakan aplikasi data mining open source berbasis Java.

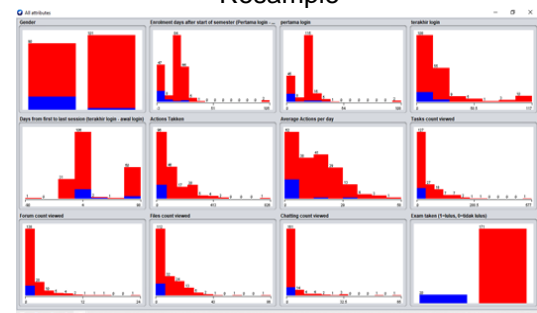
3.2. Imbalance Data

Pada penelitian ini penulis menemukan permasalahan dimana adanya imbalance data pada dataset yang penulis gunakan dalam penelitian ini, oleh karena itu untuk meningkatkan performa klasifikasi, maka penulis

melakukan resampling data dengan teknik SMOTE dan Resample.



Gambar 4. Visualize All Atribut Dataset Asli + Resample



Gambar 5. Visualize All Atribut Dataset Asli + SMOTE

3.3. Pengujian Model Klasifikasi

Pada bagian ini penulis melakukan percobaan pada Dataset Log Activity E-Learning Students dengan Menggunakan Algoritma Random Forest, Neural Network, C.45, dan AdaBoost + C4.5.

Tabel 3. Model eksperimen yang diajukan

MODEL	DATASE T	PREPROCES S	CLASSIFY
Model 1			C.45
Model 2			RF
Model 3		-	NN
Model 4			Adb+C.45
Model 5			C.45
Model 6	Asli	Resample	RF
Model 7			NN
Model 8			Adb+C.45
Model 9			C.45
Model 10		SMOTE	RF
Model 11			NN
Model 12			Adb+C.45

Pada table 3. dapat dilihat model eksperimen yang diajukan dengan melakukan perbandingan uji coba 4 algoritma, yaitu : Algoritma Random Forest, Neural Network, C.45, dan AdaBoost + C4.5 dan teknik imbalance SMOTE dan Resample. Yang

nantinya akan dicari model eksperimen yang terbaik untuk memprediksi kinerja siswa E-learning pada penelitian ini.

Tabel 4. Hasil Confusion Matrix

CLASSIFY	TP(bb)	FN(ab)	FP(ba)	TN(aa)
C.45	168	3	10	0
RF	169	2	10	0
NN	169	2	10	0
Adb+C.45	168	3	8	2
C.45	80	10	0	90
CLASSIFY	TP(bb)	FN(ab)	FP(ba)	TN(aa)
NN	79	11	0	90
Adb+C.45	81	9	0	90
C.45	163	8	13	7
RF	166	5	15	5
NN	163	8	10	10
Adb+C.45	164	7	10	10

Pada table 4. dapat dilihat perbandingan *confusion matrix* dari uji coba 4 algoritma, yaitu : Algoritma Random Forest, Neural Network, C.45, dan AdaBoost + C4.5 dan teknik imbalance SMOTE dan Resample.

Tabel 5. Perbandingan Hasil Eksperimen Pengujian Model

CLASSIFY	ROC	Kappa	Accuracy
C.45	0.55	-0.26	92.82%
RF	0.783	-0.018	93.3702
NN	0.81	-0.02	93.37%
Adb+C.45	0.82	0.24	93.92%
C.45	0.95	0.89	94.44%
RF	0.98	0.89	94.44%
NN	0.97	0.88	93.89%
Adb+C.45	0.97	0.90	95.00%
C.45	0.90	0.34	89.01%
RF	0.89	0.29	89.53%
NN	0.90	0.47	90.58%
Adb+C.45	0.91	0.49	91.10%

Dari tabel uji coba diatas diketahui Algoritma C.45 + Resample + AdaBoost yang memiliki hasil terbaik dengan *Accuracy* 95.00%, *ROC* 0.97, dan *Kappa* 0.90.

4. Kesimpulan

Penelitian ini dibuat untuk mengembangkan model prediksi kinerja siswa berdasarkan Log Activity Mahasiswa dalam E-Learning menggunakan platform moodle menggunakan

algoritma Adaptive Boosting dengan Base Learner C.45 ditambah teknik Resample untuk Imbalance dataset. Dari hasil evaluasi dan validasi, diketahui bahwa hasil uji coba algoritma Adaptive Boosting dengan Base Learner C4.5 dengan teknik Resample menghasilkan nilai performansi yang baik dengan nilai *Accuracy* 95%, *ROC* 0.97, dan *Kappa* 0.90, sehingga dapat menjadi model untuk memprediksi kinerja siswa dilihat dari log aktivitasnya menggunakan platform moodle. Untuk pengembangan pada penelitian selanjutnya dapat dilakukan penambahan jumlah data yang digunakan dan juga melakukan optimasi lainnya, seperti atribut selection menggunakan teknik PSO atau lainnya.

Referensi

- Alberth, Y., & Soepranoto, A. H. H. (2022). Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas. *Jurnal Komputer Dan Informatika*, 10(1), 31–38. <https://doi.org/10.35508/JICON.V10I1.6554>
- Aleksandrova, Y. (2019). *Predicting Students Performance In Moodle Platforms Using Machine Learning Algorithms.pdf*. 1, 177–187. <https://ideas.repec.org/a/vrn/katinf/y2019i1p177-187.html>
- Annisa, R.-, & Sasongko, A.-. (2020). Prediksi Nilai Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes. *JST (Jurnal Sains Dan Teknologi)*, 9(1), 1–10. <https://doi.org/10.23887/jst-undiksha.v9i1.19488>
- Bahad, P., & Saxena, P. (2020). *Study of AdaBoost and Gradient Boosting Algorithms for Predictive Analytics*. 235–244. https://doi.org/10.1007/978-981-15-0633-8_22
- Chango, W., Cerezo, R., & Romero, C. (2019). Predicting academic performance of university students from multi-sources data in blended learning. *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/3368691.3368694>
- He, H., Zhang, W., & Zhang, S. (2018). A novel ensemble method for credit scoring: Adaption of different imbalance ratios. *Expert Systems with Applications*, 98, 105–117. <https://doi.org/10.1016/j.eswa.2018.01.012>
- Holiver, N., Kurbatova, T., & Bondar, I. (2020). Blended learning for sustainable education: Moodle-based English for Specific Purposes teaching at Kryvyi Rih

- National University. *E3S Web of Conferences*, 166, 10006. <https://doi.org/10.1051/E3SCONF/202016610006>
- Ina, W. T., Manu, S., Odja, M., Mauboy, E., Rantelobo, K., Maggang, A., & Pella, S. (2019). Klasifikasi Tingkat Kelulusan Mahasiswa Prodi Teknik Elektro Fst Undana Menggunakan Algoritma C4.5. *Seminar Nasional Sains Dan Teknik Fst Undana (Sainstek-Iv)*, 355–361. <https://conference.undana.ac.id/sainstek/article/view/64>
- Irfan, M., Kusumaningrum, B., Yulia, Y., & Widodo, S. A. (2020). Challenges During The Pandemic: Use Of E-Learning In Mathematics Learning In Higher Education. *Infinity Journal*, 9(2), 147–158. <https://doi.org/10.22460/INFINITY.V9I2.P147-158>
- Kurniawan, Y. I. (2018). Rancang Bangun Sistem Informasi Surat Menyurat Berbasis SAAS (Software As A Service). *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, x(No. 30/E/KPT/2018), 1–8. <https://doi.org/10.25126/jtiik>
- Laila Qadrini, O., Seppewali, A., Aina, A., Studi Statistika, P., Matematika dan Ilmu Pengetahuan Alam, F., & Sulawesi Barat, U. (2021). Decision Tree dan Adaboost pada Klasifikasi Penerima Program Bantuan Sosial. *Jurnal Inovasi Penelitian*, 2(7), 1959–1966. <https://doi.org/10.47492/JIP.V2I7.1046>
- Lin, C., Tsai, C. F., & Lin, W. C. (2022). Towards hybrid over- and under-sampling combination methods for class imbalanced datasets: an experimental study. *Artificial Intelligence Review*, 56(2), 845–863. <https://doi.org/10.1007/S10462-022-10186-5/METRICS>
- Masangu, L., Jadhav, A., & Ajoodha, R. (2021). Predicting student academic performance using data mining techniques. *Advances in Science, Technology and Engineering Systems*, 6(1). <https://doi.org/10.25046/aj060117>
- Nespereira, C. G., Elhariri, E., El-Bendary, N., Vilas, A. F., & Redondo, R. P. D. (2016). Machine Learning Based Classification Approach for Predicting Students Performance in Blended Learning. *Advances in Intelligent Systems and Computing*, 407, 287–296. <https://doi.org/10.1007/978-3-319-26690-9>
- Pebrianti, L., Aulia, F., Nisa, H., & S, K. S. (2022). Implementasi Metode Adaboost untuk Mengoptimasi Klasifikasi Penyakit Diabetes dengan Algoritma Naive Bayes. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, 7(2), 122–127. <https://doi.org/10.32528/JUSTINDO.V7I2.8627>
- Putri, D. A., Hananto, B., Afrizal, S., & Pangaribuan, A. B. (2020). Prediksi Program Studi Berdasarkan Nilai Siswa Dengan Algoritma Backpropagation (Studi Kasus Sman 6 Depok Jurusan Ips). *Informatik : Jurnal Ilmu Komputer*, 15(2). <https://doi.org/10.52958/iftk.v15i2.1420>
- Raihan, M. R., Chrisnanto, Y. H., & Ningsih, A. K. (2022). Klasifikasi Penentuan Kelayakan Pinjaman Koperasi Dengan Algoritma Cart Menggunakan Algoritma Adaboost. *INFOTECH Journal*, 8(2), 74–83. <https://doi.org/https://doi.org/10.31949/info-tech.v8i2.3247>
- Shedriko, S., & Firdaus, M. (2022). Penentuan Klasifikasi Dengan Crisp-Dm Dalam Memprediksi Kelulusan Mahasiswa Pada Suatu Mata Kuliah. *Semnas Ristek (Seminar Nasional Riset Dan Inovasi Teknologi)*, 6(1). <https://doi.org/10.30998/semnasristek.v6i1.5814>
- Sudaryanto, S. N., & Sudaryanto, dan. (2022). Sintesis Fitur Density Based Feature Selection (DBFS) dan AdaBoost dengan XGBoost Untuk Meningkatkan Performa Model Prediksi. *Prosiding Seminar Sains Nasional Dan Teknologi*, 12(1), 305–313. <https://doi.org/10.36499/PSNST.V12I1.6997>