

Memprediksi Kinerja Akademik Mahasiswa menggunakan XGBoost

Laporan Tugas Akhir

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana
Informatika Universitas Muhammadiyah Malang



PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2024

LEMBAR PERSETUJUAN

Memprediksi Kinerja Akademik Mahasiswa dengan Menggunakan XGBoost

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1

Informatika Universitas Muhammadiyah Malang



LEMBAR PENGESAHAN

Memprediksi Kinerja Akademik Mahasiswa dengan Menggunakan XGBoost

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1

Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

NADIA NUR OKTAVIANI SUKMA

202010370311320

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengaji
pada tanggal 20 Maret 2024

Menyelanjui,

Dosen Pengaji 1



Dosen Pengaji 2



Wildan Suharso S.Kom., M.Kom

NIP. 10817030596PNS.

Briansyah Setio Wiyono S.Kom.,

M.Kom

NIP. 190913071987PNS.

Mengetahui,

Ketua Jurusan Informatika



Ir. Galih Wasis Wicaksono S.kom., M.Cs.

NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : NADIA NUR OKTAVIANI SUKMA

NIM : 202010370311320

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul "**Memprediksi Kinerja Akademik Mahasiswa dengan Menggunakan XGBoost**" beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing



Vinna Rahmayanti S.Si., M.Si

Malang, 20 Maret
Yang Membuat Pernyataan



NADIA NUR OKTAVIANI
SUKMA

Hadiwijaya

ABSTRACT

Predicting the academic performance of students involves comprehending, analyzing, and forecasting the diverse factors that can impact their achievements. One of the key objectives of this study is to enhance our understanding of the current academic performance of students, utilizing data sourced from a predefined objective—specifically, the grades received in the final semester exams, which were extracted from a Kaggle dataset. Various methods can be utilized for predicting student performance, and among the well-regarded approaches is regression. The employed technique is Extreme Gradient Boosting, commonly referred to as XGBoost. The outcomes derived from the XGBRegressor method reveal an MSE of 0.036, RMSE of 0.192, and MAE of 0.148. These outcomes signify that the model provides precise predictions with minimal errors, as lower error values are indicative of accurate predictions.

Keywords: *Student performance, gradient boosting, XGBoost, XGBRegressor, deep learning, machine learning*



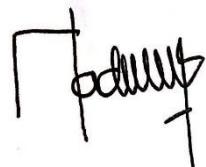
LEMBAR PERSEMBAHAN

Puji syukur kepada Allah SWT atas rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir ini. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Edi Setiadi Sukma,S.Pd. Selaku donator utama saya, tulang punggung keluarga saya, serta cinta pertama saya yang selalu mensupport, mendoakan dan berusaha untuk kesuksesan anak-anaknya. Terima kasih atas kerja keras ayah, terimakasih sudah menjadikan dan membuat Nadia ada dititik ini. Ayah harus selalu ada dan menyaksikan, menemani perjalanan Nadia sampai Nadia Benar-benar bisa membuat ayah bangga, membahagiakan ayah, karena Nadia akan selalu butuh sosok penjaga dan pelindung seperti ayah.
2. Ibu Nani Adiyana,S.Pd, selaku donatur kehidupan saya, surga saya, tempat cerita saya yang selalu mendoakan dan memberikan wejangan serta nasehat untuk saya. Terimakasih atas doa ibu yang sudah membuat nadia ada dititik ini. Ibu harus selalu ada dan menyaksikan, menemani perjalanan Nadia sampai Nadia Benar-benar bisa membuat ibu bangga, membahagiakan ibu, karena Nadia akan selalu butuh sosok ibu yang menjadi tempat Nadia untuk pulang.
3. Untuk tentunya diri saya sendiri Nadia Nur Oktaviani Sukma yang perlu diapresiasi karena sudah bertahan sejauh ini dan sampai di titik ini. Banyak sekali perjalanan kuliah yang sudah saya lewati, banyak pelajaran yang saya dapatkan, dan semua itu selalu ada hikmahnya. Terimakasih untuk diri saya sendiri sudah kuat dan bisa melewati sampai dititik ini. Saya bangga dengan diri saya sendiri.
4. Wahyu Rizky selaku rekan saya dalam menyelesaikan Jurnal tugas akhir. Orang yang semisalkan dia tidak mau saya ajak dalam jurnal ini, mungkin saya tidak akan ada dititik ini. Terimakasih sudah mau berjuang bersama dalam menyelesaikan jurnal dengan hasil yang cukup memuaskan.
5. Ibu Vinna Rahmayanti S, S.Si., M.Si selaku pembimbing pertama saya dan Bapak Christian Sri Kusuma Aditya S.Kom., M.Kom selaku pembimbing kedua saya, serta penulis pertama didalam jurnal kami. Terimakasih banyak telah membimbing saya dan Wahyu sehingga kami dapat menyelesaikan skripsi ini.

6. Bapak/Ibu Dekan Fakultas Teknik Universitas Muhammadiyah Malang.
7. Bapak/Ibu Ketua Jurusan Teknik Informatika Universitas Muhammadiyah Malang.
8. Serta Teman dekat saya, sahabat dan teman – teman saya yang tidak bisa saya sebutkan satu persatu, yang sudah membantu dan mensupport saya hingga saya dapat menyelesaikan jurnal dan ada dititik ini.

Malang, 15 Januari 2024



Nadia Nur Oktaviani Sukma



KATA PENGANTAR

Dengan memanjangkan puji syukur kehadiran Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul :

“Memprediksi Prestasi Akademik Siswa menggunakan XGBoost”

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi algoritma dan deskripsi dataset yang digunakan, pra-pemrosesan yang dipilih, evaluasi dan hasil pengujian pada dataset yang telah ditentukan.

Peneliti menyadari sepenuhnya bahwa dalam penulisan tugas akhir ini masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 15 Januari 2024



Nadia Nur Oktaviani Sukma



DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
LEMBAR PERSEMBAHAN	vii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xi
DAFTAR TABEL	xi
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Batasan Penelitian.....	4
TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Google Colaboratory.....	6
2.3 Bahasa Pemrograman Python	6
2.4 <i>Student Performance</i>	6
2.5 Machine Learning	7
2.5.1 eXtreme Gradient Boosting (XGBoost) Regressor.....	7
METODOLOGI PENELITIAN	9
3.1 Objek Penelitian.....	9
3.2 Rancangan Penelitian.....	9
3.3 Dataset.....	10
3. 4 Preprocessing Data.....	10
3.4.1 Missing Value	11
3.4.2 Data duplicate	11
3.4.3 Encoding	12
3.4.4 Feature Selection PCA (Principal Component Analysis)	12
3.5 Model <i>eXtreme Gradient Boosting</i> (XGBoost) Regressor	12
3.6 Evaluasi dan Pengujian	13
HASIL DAN PEMBAHASAN	15
PENUTUP	18

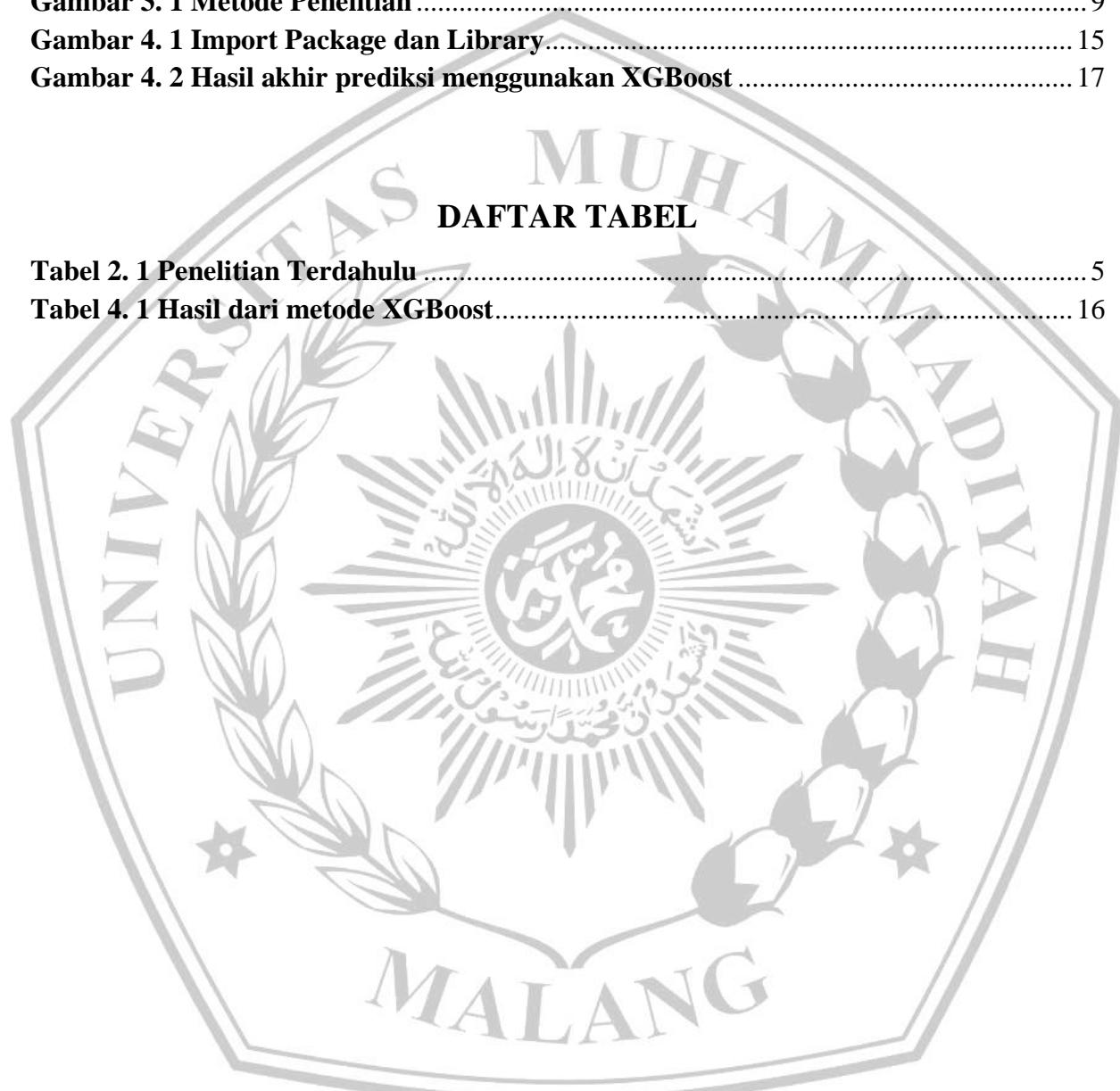
5.1 Kesimpulan	18
5.2 Saran	18
DAFTAR PUSTAKA	19

DAFTAR GAMBAR

Gambar 2. 1 Struktur model eXtreme Gradient Boosting (XGBoost) Regressor	8
Gambar 3. 1 Metode Penelitian	9
Gambar 4. 1 Import Package dan Library.....	15
Gambar 4. 2 Hasil akhir prediksi menggunakan XGBoost	17

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	5
Tabel 4. 1 Hasil dari metode XGBoost.....	16



DAFTAR PUSTAKA

- [1] R. V. Adams and E. Blair, "Impact of Time Management Behaviors on Undergraduate Engineering Students' Performance," *SAGE Open*, vol. 9, no. 1, 2019, doi: 10.1177/2158244018824506.
- [2] D. Buenaño-Fernández, D. Gil, and S. Luján-Mora, "Application of machine learning in predicting performance for computer engineering students: A case study," *Sustain.*, vol. 11, no. 10, pp. 1–18, 2019, doi: 10.3390/su11102833.
- [3] A. Namoun and A. Alshanqiti, "Predicting student performance using data mining and learning analytics techniques: A systematic literature review," *Applied Sciences (Switzerland)*, vol. 11, no. 1. MDPI AG, pp. 1–28, Jan. 01, 2021. doi: 10.3390/app11010237.
- [4] S. Qu, K. Li, B. Wu, X. Zhang, and K. Zhu, "Predicting student performance and deficiency in mastering knowledge points in MOOCs using multi-task learning," *Entropy*, vol. 21, no. 12, 2019, doi: 10.3390/e21121216.
- [5] N. Amelia, A. G. Abdullah, and Y. Mulyadi, "Meta-analysis of student performance assessment using fuzzy logic," *Indones. J. Sci. Technol.*, vol. 4, no. 1, pp. 74–88, 2019, doi: 10.17509/ijost.v4i1.15804.
- [6] M. Tsiakkaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer learning from deep neural networks for predicting student performance," *Appl. Sci.*, vol. 10, no. 6, 2020, doi: 10.3390/app10062145.
- [7] F. Ofori, E. Maina, and R. Gitonga, "Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review," *J. Inf. Technol.*, vol. 4, no. 1, pp. 2616–3573, 2020, [Online]. Available: <https://stratfordjournals.org/journals/index.php/Journal-of-Information-and-Technol/article/view/480>
- [8] L. Kemper, G. Vorhoff, and B. U. Wigger, "Predicting student dropout: A machine learning approach," *Eur. J. High. Educ.*, vol. 10, no. 1, pp. 28–47, 2020, doi: 10.1080/21568235.2020.1718520.
- [9] T. T. Dien, S. H. Luu, N. Thanh-Hai, and N. Thai-Nghe, "Deep Learning with Data

Transformation and Factor Analysis for Student Performance Prediction," 2020. [Online]. Available: www.ijacsa.thesai.org

- [10] J. L. Rastrollo-Guerrero, J. A. Gómez-Pulido, and A. Durán-Domínguez, "Analyzing and predicting students' performance by means of machine learning: A review," *Applied Sciences (Switzerland)*, vol. 10, no. 3. MDPI AG, Feb. 01, 2020. doi: 10.3390/app10031042.
- [11] P. Will, W. F. Bischof, and A. Kingstone, "The impact of classroom seating location and computer use on student academic performance," *PLoS One*, vol. 15, no. 8 August 2020, Aug. 2020, doi: 10.1371/journal.pone.0236131.
- [12] F. Qiu *et al.*, "Predicting students' performance in e-learning using learning process and behaviour data," *Sci. Rep.*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-021-03867-8.
- [13] A. O. Oyedele, A. M. Salami, O. Folorunsho, and O. R. Abolade, "Analysis and Prediction of Student Academic Performance Using Machine Learning," *JITCE (Journal Inf. Technol. Comput. Eng.)*, vol. 4, no. 01, pp. 10–15, Mar. 2020, doi: 10.25077/jitce.4.01.10-15.2020.
- [14] A. U. Khasanah and H. Harwati, "Educational data mining techniques approach to predict student's performance," *Int. J. Inf. Educ. Technol.*, vol. 9, no. 2, pp. 115–118, Feb. 2019, doi: 10.18178/ijiet.2019.9.2.1184.
- [15] A. Gavrikov, Y. Malyshkin, and F. Ratnikov, "Energy reconstruction for large liquid scintillator detectors with machine learning techniques: aggregated features approach," *Eur. Phys. J. C*, vol. 82, no. 11, pp. 1–12, 2022, doi: 10.1140/epjc/s10052-022-11004-6.
- [16] A. T. Tunkiel, D. Sui, and T. Wiktorski, "Reference dataset for rate of penetration benchmarking," *J. Pet. Sci. Eng.*, vol. 196, no. September 2020, p. 108069, 2021, doi: 10.1016/j.petrol.2020.108069.
- [17] J. Ge *et al.*, "Prediction of Greenhouse Tomato Crop Evapotranspiration Using XGBoost Machine Learning Model," *Plants*, vol. 11, no. 15, pp. 1–17, 2022, doi: 10.3390/plants11151923.
- [18] T. Carneiro, R. V. M. Da Nobrega, T. Nepomuceno, G. Bin Bian, V. H. C. De Albuquerque, and P. P. R. Filho, "Performance Analysis of Google Colaboratory as a Tool for Accelerating Deep Learning Applications," *IEEE Access*, vol. 6, pp. 61677–61685, 2018, doi: 10.1109/ACCESS.2018.2874767.
- [19] S. K. R., "Python -The Fastest Growing Programming Language," *Int. Res. J. Eng. Technol.*, vol. 4, no. 12, pp. 354–357, 2017, [Online]. Available: www.irjet.net

- [20] S. Saharuddin and M. Wisnu Prihatmono, “Pengenalan Dan Pelatihan Dasar Bahasa Pemrograman Python Pada Siswa/I Sma Negeri 3 Makassar,” *SELAPARANG J. Pengabdi. Masy. Berkemajuan*, vol. 6, no. 4, p. 2233, 2022, doi: 10.31764/jpmb.v6i4.10569.
- [21] S. Badillo *et al.*, “An Introduction to Machine Learning,” vol. 00, no. 0, pp. 1–15, 2020, doi: 10.1002/cpt.1796.
- [22] K. Qadeer, W. U. Rehman, A. M. Sheri, I. Park, H. K. Kim, and M. Jeon, “A Long Short-Term Memory (LSTM) network for hourly estimation of PM2.5 concentration in two cities of South Korea,” *Appl. Sci.*, vol. 10, no. 11, 2020, doi: 10.3390/app10113984.
- [23] M. Ahtesham, N. Z. Bawany, and K. Fatima, “House Price Prediction using Machine Learning Algorithm - The Case of Karachi City, Pakistan,” *Proc. - 2020 21st Int. Arab Conf. Inf. Technol. ACIT 2020*, pp. 7–11, 2020, doi: 10.1109/ACIT50332.2020.9300074.
- [24] D. P. M. Abellana and D. M. Lao, “A new univariate feature selection algorithm based on the best–worst multi-attribute decision-making method,” *Decis. Anal. J.*, vol. 7, no. May, p. 100240, 2023, doi: 10.1016/j.dajour.2023.100240.
- [25] V. Mistry, A. K. Mishra, and N. Ahmed, “Machine Learning Use Case in Indian Agriculture: Predictive Analysis of Bihar Agriculture Data to Forecast Crop Yield,” *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 11, no. 1, pp. 1004–1009, 2023, doi: 10.22214/ijraset.2023.48709.
- [26] S. Li *et al.*, “Automated high-throughput genome editing platform with an AI learning in situ prediction model,” *Nat. Commun.*, vol. 13, no. 1, pp. 1–11, 2022, doi: 10.1038/s41467-022-35056-0.
- [27] M. Yang, B. Zhang, Y. Chen, X. Xin, K. Lee, and B. Chen, “Impact of microplastics on oil dispersion efficiency in the marine environment,” *Sustain.*, vol. 13, no. 24, pp. 1–13, 2021, doi: 10.3390/su132413752.

**FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR**

Nama Mahasiswa : Nadia Nur Oktaviani Sukma

NIM : 202010370311320

Judul TA : Memprediksi Kinerja Akademik Mahasiswa menggunakan
XGBoost**Hasil Cek Plagiarisme dengan Turnitin**

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	2 %
2.	Bab 2 – Daftar Pustaka	25 %	6 %
3.	Bab 3 – Analisis dan Perancangan	25 %	4 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	4 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	4 %
6.	Makalah Tugas Akhir	20%	4 %

*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,**Pemeriksa (Staff TU)**

 (.....)
**Kampus I**
 Jl. Bandung 1 Malang, Jawa Timur
 P. +62 341 551 253 (Hunting)
 F. +62 341 460 435
Kampus II
 Jl. Bendungan Sutama No.158 Malang, Jawa Timur
 P. +62 341 551 149 (Hunting)
 F. +62 341 582 080
Kampus III
 Jl. Raya Tlogomas No.248 Malang, Jawa Timur
 P. +62 341 464 318 (Hunting)
 F. +62 341 460 435
 E. webmaster@umm.ac.id