

**Deteksi Bias dalam Model Machine Learning untuk Prediksi
Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning
Environment**

Laporan Tugas Akhir

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



Deva Putra Setya Pratama

202010370311212

Bidang Minat

Data Science

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH MALANG
2024**

LEMBAR PERSETUJUAN

**Deteksi Bias dalam Model Machine Learning untuk Prediksi
Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning
Environment**

TUGAS AKHIR

**Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang**

Menyetujui,

Malang, 16 Juli 2024

Dosen Pembimbing 1



Setio Basuki MT., Ph.D.

NIP. 10809070477PNS.

LEMBAR PENGESAHAN

Deteksi Bias dalam Model Machine Learning untuk Prediksi Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning Environment

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

Deva Putra Setya Pratama

202010370311212

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian
pada tanggal 16 Juli 2024

Menyetujui,

Dosen Penguji 1



Vinna Rahmayanti S S.Si., M.Si

NIP. 180306071990PNS.

Dosen Penguji 2



Christian Sri Kusuma Aditya

S.Kom., M.Kom

NIP. 180327021991PNS.

Mengetahui,

Ketua Jurusan Informatika



Galih Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Deva Putra Setya Pratama

NIM : 202010370311212

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul “**Deteksi Bias dalam Model Machine Learning untuk Prediksi Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning Environment**” 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



Setio Basuki MT., Ph.D.

Malang, 16 Juli 2024
Yang Membuat Pernyataan



Deva Putra Setya Pratama

ABSTRAK

Revolusi digital yang cepat dalam pendidikan telah menempatkan *Virtual Learning Environment* (VLE) sebagai elemen penting dalam mengembangkan paradigma pembelajaran, yang sangat terlihat selama pandemi COVID-19. Penelitian ini menyelidiki dampak aktivitas siswa berbasis VLE dalam memprediksi keberhasilan akademik dan mengatasi bias dalam model machine learning yang digunakan untuk prediksi ini. Menggunakan *Open University Learning Analytics Dataset* (OULAD), penelitian ini mengintegrasikan teknik prapemrosesan data, pemilihan fitur, dan transformasi data untuk mengembangkan dataset yang komprehensif. Model Random Forest digunakan untuk memprediksi hasil kelulusan siswa, yang dikategorikan menjadi kelas "*pass*", "*fail*", dan "*distinction*". Kinerja model dievaluasi menggunakan metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score, serta matriks kebingungan. Deteksi bias dilakukan menggunakan alat DALEX, dengan fokus pada atribut terlindungi seperti usia, jenis kelamin, dan disabilitas untuk memastikan keadilan. Hasilnya mengungkapkan akurasi model yang tinggi tetapi menyoroti adanya bias yang signifikan dalam beberapa kelompok demografis. Penelitian ini berkontribusi pada diskursus berkelanjutan tentang memastikan penerapan machine learning yang etis dan adil dalam pengaturan pendidikan dengan mengusulkan metode untuk meningkatkan kesetaraan dan transparansi model prediktif.

Kata Kunci: Lingkungan Pembelajaran Virtual, Pembelajaran Mesin, Deteksi Bias, DALEX, Open University Learning Analytics Dataset, Random Forest, Penggalan Data Pendidikan

ABSTRACT

The rapid digital revolution in education has positioned Virtual Learning Environments (VLEs) as critical to evolving learning paradigms, particularly highlighted during the COVID-19 pandemic. This research investigates the impact of VLE-based student activity on predicting academic success and addresses the biases in machine learning models used for these predictions. Using the Open University Learning Analytics Dataset (OULAD), this study integrates data preprocessing techniques, feature selection, and data transformation to develop a comprehensive dataset. A Random Forest model is employed to predict student graduation outcomes, categorized into "pass", "fail", and "distinction" classes. The model's performance is evaluated using classification metrics such as accuracy, precision, recall, and F1-score, alongside confusion matrices. Bias detection is conducted using the DALEX tool, focusing on protected attributes like age, gender, and disability to ensure fairness. The results reveal high model accuracy but highlight significant bias in some demographic groups. This study contributes to the ongoing discourse on ensuring ethical and fair machine learning applications in educational settings by proposing methods to enhance the equity and transparency of predictive models.

Keywords: *Virtual Learning Environment, Machine Learning, Bias Detection, DALEX, Open University Learning Analytics Dataset, Random Forest, Educational Data Mining*

KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT karena atas rahmat dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul "**Deteksi Bias dalam Model Machine Learning untuk Prediksi Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning Environment**". Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Universitas Muhammadiyah Malang. Penulisan skripsi ini tidak lepas dari bantuan, dukungan, dan bimbingan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT sebagai pencipta seluruh alam, telah memberikan kesehatan dan kelancaran kepada penulis dalam mengerjakan penelitian ini.
2. Kedua orang tua penulis Bapak Sutiono dan Ibu Nanik Sri Wahyuni yang selalu memberikan doa, dukungan, biaya, dan kasih sayang tanpa henti.
3. Bapak Setio Basuki MT., Ph.D., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan motivasi yang sangat berarti selama proses penyusunan skripsi ini.
4. Ibu Vinna Rahmayanti S S.Si., M.Si dan Bapak Christian Sri Kusuma Aditya S.Kom., M.Kom, selaku dosen penguji yang telah memberikan kritik, saran, dan masukan yang sangat berharga untuk perbaikan skripsi ini.
5. Seluruh dosen dan staf administrasi di Program Studi Informatika Universitas Muhammadiyah Malang yang telah membantu dalam berbagai keperluan akademik dan administratif.
6. Teman-teman seperjuangan yang telah memberikan semangat dan kebersamaan selama masa studi.
7. Saya pribadi yang telah berhasil menyelesaikan studi.

Demikian kata pengantar ini penulis sampaikan. Penulis berharap skripsi ini bermanfaat bagi pembaca dan pengembangan ilmu pengetahuan di bidang ini.

Malang, 16 Juli 2024



Deva Putra Setya Pratama

DAFTAR ISI

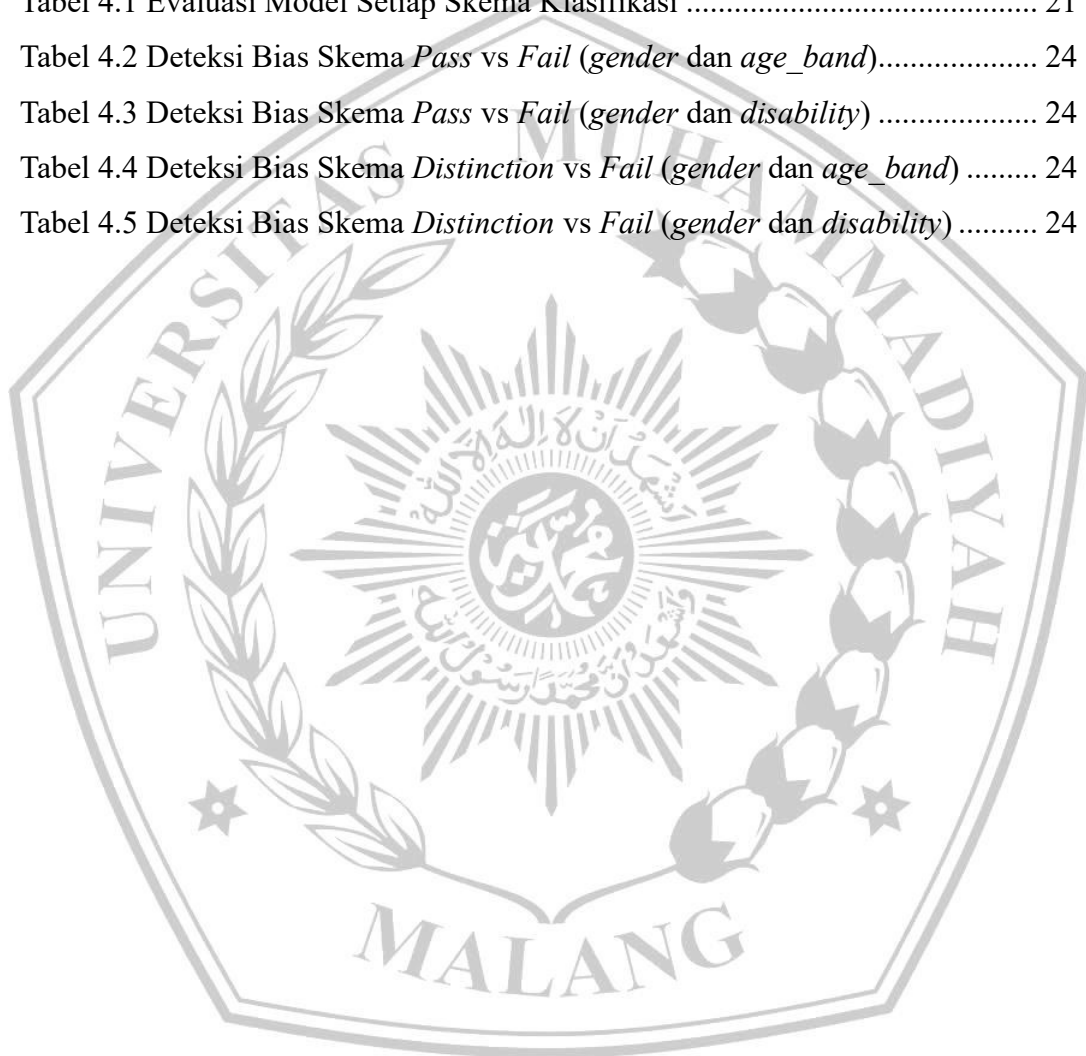
LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	4
1.3 Tujuan.....	5
1.4 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu.....	6
2.2 Virtual Learning Environment (VLE).....	9
2.3 Machine Learning	9
2.4 Random Forest	10
2.5 Dalex	10
BAB III METODOLOGI PENELITIAN	11
3.1 Rancangan Penelitian.....	11
3.2 Dataset.....	11
3.3 Pemrosesan Data	12
3.4 Klasifikasi	17
3.5 Evaluasi	19
3.6 Deteksi Bias	20
BAB IV HASIL DAN PEMBAHASAN.....	21
4.1 Evaluasi Model.....	21
4.2 Deteksi Bias	23
BAB V KESIMPULAN.....	27

DAFTAR PUSTAKA..... 28
LAMPIRAN..... 30



DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu Terkait Klasifikasi dan Pengolahan Data	6
Tabel 2.2 Penelitian Terdahulu Terkait Deteksi Bias	8
Tabel 3.1 Atribut OULAD dan Detail Pemrosesan	13
Tabel 3.2 Rancangan Entitas Baru dari Pemrosesan Data	15
Tabel 3.3 Parameter Model <i>Random Forest</i>	18
Tabel 4.1 Evaluasi Model Setiap Skema Klasifikasi	21
Tabel 4.2 Deteksi Bias Skema <i>Pass vs Fail (gender dan age_band)</i>	24
Tabel 4.3 Deteksi Bias Skema <i>Pass vs Fail (gender dan disability)</i>	24
Tabel 4.4 Deteksi Bias Skema <i>Distinction vs Fail (gender dan age_band)</i>	24
Tabel 4.5 Deteksi Bias Skema <i>Distinction vs Fail (gender dan disability)</i>	24



DAFTAR GAMBAR

Gambar 3.1 Rancangan Penelitian Deteksi Bias.....	11
Gambar 3.2 Skema Database OULAD	12
Gambar 3.3 Distribusi Pass vs Fail	19
Gambar 3.4 Distribusi Distinction vs Fail	19
Gambar 4.1 Confusion Matrix Pass vs Fail	22
Gambar 4.2 Confusion Matrix Distinction vs Fail.....	22



DAFTAR LAMPIRAN

Lampiran 1. Sampel Dataset Final	30
Lampiran 2. Sertifikat Plagiasi.....	32



DAFTAR PUSTAKA

- [1] A. H. A Rashid, N. A. Shukor, Z. Tasir, and S. N. Kew, "Teachers' perceptions and readiness toward the implementation of virtual learning environment," *IJERE (International Journal of Evaluation and Research in Education)*, vol. 10, no. 1, pp. 209–214, 2021, doi: 10.11591/ijere.v10i1.21014.
- [2] R. Yusny and G. I. Yasa, "Mengembangkan (Pembelajaran) blended learning dengan sistem lingkungan pembelajaran virtual (Vle) Di ptkin," *Jurnal Ilmiah Islam Futura*, vol. 19 no. 1, pp. 103–127, 2019, doi: 10.22373/jiif.v19i1.3707.
- [3] Y. A. Hambali, R. R. Putra, and Wahyudin, "Implementasi Metaverse menggunakan Aplikasi gather town untuk Pendidikan Jarak Jauh dengan Pendekatan Virtual Learning Environment," *INFORMATION SYSTEM FOR EDUCATORS AND PROFESSIONALS : Journal of Information System*, vol. 7, no.2, pp. 163–172, 2023, doi: 10.51211/isbi.v7i2.2039.
- [4] F. N. Andryas, A. Irmahayu, and N. Bustamam, "Virtual learning environment and learning approach among pre-clinical undergraduate medical students during COVID-19 pandemic," *The Indonesian Journal of Medical Education*, vol. 11, no. 1, pp. 10–21, 2022, doi: 10.22146/jpki.63975.
- [5] P. A. Petare, M. Shamim, T. Gupta, R. Verma, and G. Singh, "Exploring the impact of virtual learning environments on student engagement and academic achievement," *Journal of Survey in Fisheries Sciences*, vol. 10, no. 1S, pp. 5912–5923, 2023, doi: 10.13140/RG.2.2.23223.91040.
- [6] B. Mandasari, "The impact of online learning toward students' academic performance on business correspondence course," *EDUTECH (Journal of Education and Technology)*, vol. 4, no. 1, pp. 98–110, 2020, doi: 10.29062/edu.v4i1.74.
- [7] P. Muljana and T. Luo, "Factors contributing to student retention in online learning and recommended strategies for improvement: A systematic literature review," *Journal of Information Technology Education: Research*, vol. 18, pp. 19–57, 2019, doi: 10.28945/4182.
- [8] A. Al-Azawei and M. A. Al-Masoudy, "Predicting learners' performance in virtual learning environment (VLE) based on demographic, behavioral and engagement antecedents," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 09, pp. 60–75, 2020, doi: 10.3991/ijet.v15i09.12691.
- [9] M. S. Ahmad, A. H. Asad, and A. Mohammed, "A machine learning based approach for student performance evaluation in educational data mining," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2021, pp. 187–192, doi: 10.1109/miucc52538.2021.9447602.
- [10] N. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," *Computers & Education*, vol. 143, no. 103676, pp. 1–18, 2020, doi: 10.1016/j.compedu.2019.103676.
- [11] H. Waheed, S. Hassan, N. R. Aljohani, J. Hardman, S. Alelyani, and R. Nawaz, "Predicting academic performance of students from VLE big data using deep learning models," *Computers in Human Behavior*, vol. 104, no. 106189, 2020, doi: 10.1016/j.chb.2019.106189.
- [12] S. Rizvi, B. Rienties, and S. A. Khoja, "The role of demographics in online learning; A decision tree based approach," *Computers & Education*, vol. 137, pp. 32–47, 2019, doi: 10.1016/j.compedu.2019.04.001.

- [13] A. Al-Zawqari, D. Peumans, and G. Vandersteen, "A flexible feature selection approach for predicting students' academic performance in online courses," *Computers and Education: Artificial Intelligence*, vol. 3, no. 100103, 2022, doi: 10.1016/j.caeai.2022.100103.
- [14] R. S. Baker and A. Hawn, "Algorithmic bias in education," *International Journal of Artificial Intelligence in Education*, vol. 32, no. 4, pp. 1052–1092, 2021, doi: 10.35542/osf.io/pbmvz.
- [15] R. Fu, Y. Huang, and P. V. Singh, "AI and algorithmic bias: Source, detection, mitigation and implications," *SSRN Electronic Journal*, vol. 65, pp. 39–63, 2020, doi: 10.2139/ssrn.3681517.
- [16] K. Cachel, E. Rundensteiner, and L. Harrison, "MANI-rank: Multiple attribute and intersectional group fairness for consensus ranking," in *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2022, pp. 1124–1137, doi: 10.1109/icde53745.2022.00089.
- [17] S. Alelyani, "Detection and evaluation of machine learning bias," *Applied Sciences*, vol. 11, no. 14, pp. 6271, 2021, doi: 10.3390/app11146271.
- [18] J. Gardner, C. Brooks, and R. Baker, "Evaluating the fairness of predictive student models through slicing analysis," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 2019, pp. 225–234, doi: 10.1145/3303772.3303791.
- [19] S. Riazy, K. Simbeck, and V. Schreck, "Fairness in learning analytics: Student at-risk prediction in virtual learning environments," in *Proceedings of the 12th International Conference on Computer Supported Education*, 2020, pp. 15–25, doi: 10.5220/0009324100150025.
- [20] M. Verger, S. Lalle, F. Bouchet, and V. Luengo, "Is your model 'MADD'? A novel metric to evaluate algorithmic fairness for predictive student models," in *Proceedings of the 16th International Conference on Educational Data Mining (EDM 2023)*, 2023, doi: 10.5281/zenodo.8115786.
- [21] S. Riazy and K. Simbeck, "Predictive algorithms in learning analytics and their fairness," *Lecture Notes in Informatics (LNI) - Proceedings*, 2019, pp. 223–228, doi: 10.18420/delfi2019_305.
- [22] J. Wiśniewski and P. Biecek, "Fairmodels: A flexible tool for bias detection, visualization, and mitigation in binary classification models," *The R Journal*, vol. 14, no. 1, pp. 227–243, 2022, doi: 10.32614/rj-2022-019.
- [23] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open University Learning Analytics dataset," *Sci. Data*, vol. 4, No. 170171, pp. 1–8, 2017, doi: 10.1038/sdata.2017.171.
- [24] H. Baniecki, W. Kretowicz, P. Piatyszek, J. Wisniewski, and P. Biecek, "dalex: Responsible Machine Learning with Interactive," *Journal of Machine Learning Research*, vol. 22, pp. 1–7, 2021, doi: 10.48550/arXiv.2012.14406.
- [25] H. A. Mengash, "Using data mining techniques to predict student performance to support decision making in university admission systems," *IEEE Access*, vol. 8, pp. 55462–55470, 2020, doi: 10.1109/ACCESS.2020.2981905.
- [26] S. Alija, E. Beqiri, A. S. Gaafar, and A. K. Hamoud, "Predicting students performance using supervised machine learning based on imbalanced dataset and wrapper feature selection," *Informatica*, vol. 47, no. 1, pp. 11–19, 2023, doi: 10.31449/inf.v47i1.4519.



UNIVERSITAS
MUHAMMADIYAH
MALANG

FAKULTAS TEKNIK

INFORMATIKA

informatika.umm.ac.id | informatika@umm.ac.id



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Deva Putra Setya Pratama
 NIM : 202010370311212
 Judul TA : Deteksi Bias dalam Model Machine Learning untuk Prediksi Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning Environment

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 %	5 %
3.	Bab 3 – Analisis dan Perancangan	25 %	6 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	4 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	4 %
6.	Makalah Tugas Akhir	20%	0 %

*) 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 Sutarni No 188 Malang, Jawa Timur
 P +62 341 551 149 (Hunting)
 F +62 341 582 060

Kampus III
 Jl. Raya Tlogomas No 248 Malang, Jawa Timur
 P +62 341 464 518 (Hunting)
 F +62 341 460 435
 E. webmaster@umm.ac.id