

**PENERAPAN MACHINE LEARNING FAIRNESS PADA DATA  
TABULAR : STUDI KASUS TERHADAP DATA MEDIS**

**Laporan Tugas Akhir**

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



Muhamad Minan Nur Mu'alim

202110370311197

**Bidang Minat**

Data Science

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

**2024/2025**

## LEMBAR PERSETUJUAN

**PENERAPAN MACHINE LEARNING FAIRNESS PADA DATA  
TABULAR : STUDI KASUS TERHADAP DATA MEDIS**

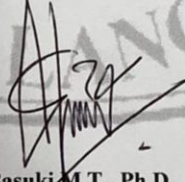
**TUGAS AKHIR**

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

Menyetujui,

Malang, 14 Mei 2025

Dosen Pembimbing 1

  
**Setio Basuki M.T., Ph.D.**  
**NIP. 10809070477PNS.**

## LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

**NAMA : MUHAMAD MINAN NUR MU'ALIM**

**NIM : 202110370311197**

**FAK./JUR. : INFORMATIKA**

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul "PENERAPAN MACHINE LEARNING FAIRNESS PADA DATA TABULAR : STUDI KASUS TERHADAP DATA MEDIS" beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali kutipan yang telah disebutkan sumbernya.

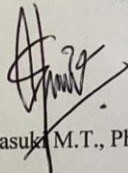
Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,

Malang, 18 Mei 2025

Dosen Pembimbing

Yang Membuat Pernyataan



Setio Basuki M.T., Ph.D.



Muhamad Minan Nur Mu'alim

# LEMBAR PENGESAHAN

## PENERAPAN MACHINE LEARNING FAIRNESS PADA DATA TABULAR : STUDI KASUS TERHADAP DATA MEDIS

### TUGAS AKHIR

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

Disusun Oleh :

**MUHAMAD MINAN NUR MU'ALIM**

**202110370311197**

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis pengujian  
pada tanggal 15 Juni 2025

Menyetujui,

Dosen Penguji 1

Dosen Penguji 2

Christian Sri Kusuma Aditya S.Kom.,

Ir. Wildan Suharso S.Kom., M.Kom

M.Kom

NIP. 10817030596PNS.

NIP. 180327021991PNS.

Mengetahui,  
Ketua Jurusan Informatika



Dr. Cahyo Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.

## ABSTRAK

Diabetes merupakan penyakit kronis yang jumlah penderitanya terus mengalami peningkatan, diperkirakan bisa mencapai berjuta-juta kasus di masa yang akan mendatang. Penelitian ini menganalisis hubungan pola hidup dengan diabetes menggunakan dataset "CDC Diabetes Health Indicators" melalui prapemrosesan dan transformasi data untuk mengoptimalkan model machine learning. Algoritma Machine Learning seperti Decision Tree, Random Forest, Logistic Regression, dan Naive Bayes digunakan untuk mengklasifikasikan individu dalam kategori "*no-diabetes*", "*pre-diabetes*", atau "*diabetes*". Kinerja model dievaluasi menggunakan metrik klasifikasi seperti akurasi, presisi, recall, F1-score, AUC, dan confusion matrix. Deteksi bias dilakukan menggunakan library DALEX dengan fokus pada atribut terlindungi seperti usia dan jenis kelamin. Hasilnya mengungkapkan model algoritma berbasis probabilitas biner seperti Gaussian Naive Bayes dianggap paling adil karena memperlakukan semua kelompok secara hampir sama, sehingga tidak ada kelompok tertentu yang terlalu diuntungkan atau dirugikan. Namun, model ini terkadang tidak seakurat model lain dalam memprediksi apakah seseorang mengidap diabetes atau tidak. Sebaliknya, model algoritma berbasis pohon seperti Random Forest memang sedikit kurang adil dibanding Naive Bayes, tapi kemampuannya dalam memprediksi benar dan salah lebih tinggi. Artinya, Random Forest bisa memberi hasil yang lebih akurat, meskipun keadilannya antar kelompok tidak sebaik Naive Bayes. Berdasarkan hasil tersebut menegaskan pentingnya penerapan machine learning yang etis dan adil dalam diagnosa penyakit, termasuk upaya meningkatkan kesetaraan dan transparansi model prediktif.

**Kata Kunci :** Diabetes, Pembelajaran Mesin, Keadilan Pembelajaran Mesin, Deteksi Bias, DALEX.

## ABSTRACT

Diabetes is a chronic disease whose prevalence continues to rise, with projections indicating that the number of cases could reach millions in the future. This research analyzes the relationship between lifestyle and diabetes using the "CDC Diabetes Health Indicators" dataset through data preprocessing and transformation to optimize machine learning models. Machine Learning algorithms such as Decision Tree, Random Forest, Logistic Regression, and Naive Bayes are used to classify individuals into the categories of "*no diabetes*", "*pre-diabetes*", or "*diabetes*". The model's performance is evaluated using classification metrics such as accuracy, precision, recall, F1-score, AUC, and confusion matrix. Bias detection was conducted using the DALEX library with a focus on protected attributes such as age and gender. The results reveal that binary probability-based algorithm models like Gaussian Naive Bayes are considered the most fair because they treat all groups almost equally, so no particular group is overly advantaged or disadvantaged. However, this model is sometimes not as accurate as other models in predicting whether someone has diabetes or not. On the other hand, tree-based algorithm models like Random Forest are indeed slightly less fair compared to Naive Bayes, but their ability to predict correctly and incorrectly is higher. This means that Random Forest can provide more accurate results, although its fairness among groups is not as good as Naive Bayes. Based on these results, this emphasizes the importance of applying ethical and fair machine learning in disease diagnosis, including efforts to enhance equity and transparency in predictive models.

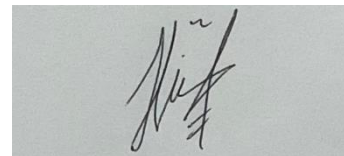
**Keywords:** Diabetes, Machine Learning, Machine Learning Fairness, Bias Detection, DALEX.

## 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. Allah SWT yang senantiasa memberikan kehidupan, kesehatan, serta kemudahan dalam setiap langkah yang saya tempuh.
2. Kedua orang tua penulis Bapak Tukimin dan Ibu Zumaroh yang senantiasa memberikan dorongan, semangat, kasih sayang, doa, dan dukungan sepenuhnya kepada penulis dalam mengerjakan penelitian ini.
3. Kedua adik penulis Ahmad Raihan Nur Riyadhoh dan Muhammad Asfahani Irihash Nuril Ilmi yang selalu memberikan semangat, dukungan, kasih sayang, dan doa kepada penulis dalam mengerjakan penelitian ini
4. Bapak Setio Basuki, MT., Ph.D. selaku pembimbing tugas akhir yang telah meluangkan waktu untuk membimbing penulis sehingga dapat menyelesaikan penelitian tugas akhir ini.
5. Bapak/Ibu Dekan Fakultas Teknik Universitas Muhammadiyah Malang.
6. Bapak Ir. Galih Wasis Wicaksono, S.Kom., M.Cs., selaku Ketua Jurusan Teknik Informatika Muhammadiyah Malang.
7. Seluruh Dosen beserta Staff Program Studi Informatika Universitas Muhammadiyah Malang yang telah memberikan ilmu dan pengalaman selama perkuliahan.
8. Kepada teman-teman yang telah memberikan dukungan tanpa henti, semangat yang tiada habisnya, berbagai masukan berharga, serta menjadi tempat berbagi dalam suka maupun duka.
9. Saya pribadi yang telah berhasil menyelesaikan studi.

Malang, 08 Agustus 2025



Muhamad Minan Nur Mu'alim

## KATA PENGANTAR

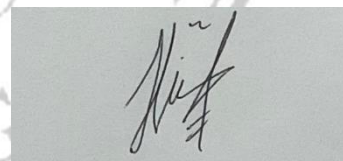
Dengan memanjatkan puji syukur kehadirat Allah SWT. Atas limpahan rahmat dan hidayah-NYA sehingga peneliti dapat menyelesaikan tugas akhir yang berjudul :

### **”PENERAPAN MACHINE LEARNING FAIRNESS PADA DATA TABULAR : STUDI KASUS TERHADAP DATA MEDIS”**

Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi latar belakang, metode penelitian, serta hasil dan pembahasan yang disimpulkan berdasarkan proses yang telah dilalui oleh peneliti.

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 Mei 2025



Muhamad Minan Nur Mu'alim

## DAFTAR ISI

LEMBAR PERSETUJUAN.....	ii
LEMBAR PERNYATAAN .....	iii
LEMBAR PENGESAHAN .....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
LEMBAR PERSEMBAHAN .....	vii
KATA PENGANTAR .....	viii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xi
DAFTAR TABEL.....	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian .....	4
1.4 Batasan Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Studi Literatur .....	5
2.2 Diabetes.....	6
2.3 Pre Processing.....	6
2.3.1 Pengecekan Nilai Hilang.....	6
2.3.2 Encoding Data.....	7
2.3.3 Normalisasi Data.....	7
2.3.4 Split data .....	7
2.4 Machine Learning .....	8
2.4.1 Decision Tree .....	9
2.4.2 Random Forest.....	9
2.4.3 Logistic Regresion .....	10
2.4.4 Naive Bayes .....	10
2.5 Dalex .....	11
2.6 Machine Learning Fairness.....	12
2.7 Evaluasi.....	13
2.7.1 Evaluasi Model.....	13

2.7.2 Deteksi Bias .....	15
BAB III METODOLOGI PENELITIAN.....	21
3.1 Rancangan Penelitian .....	21
3.2 Dataset.....	22
3.3 Pre Processing .....	26
3.3.1 Pengecekan Nilai Hilang.....	26
3.3.2 Encoding Data.....	27
3.3.3 Normalisasi Data.....	27
3.3.4 Split data .....	27
3.4 Machine Learning .....	28
3.4.1 Decision Tree .....	28
3.4.2 Logistic Regression.....	28
3.4.3 Random Forest.....	29
3.4.4 Naive Bayes .....	29
3.5 Implementasi Library Dalex .....	29
3.6 Machine Learning Fairness.....	30
3.7 Evaluasi.....	31
3.7.1 Evaluasi Model.....	31
3.7.2 Deteksi Bias .....	31
BAB IV HASIL DAN PEMBAHASAN .....	32
4.1 Evaluasi Model.....	32
4.1.1 Dataset 1 .....	32
4.1.2 Dataset 2.....	36
4.1.3 Dataset 3.....	40
4.2 Deteksi Bias .....	43
4.2.1 DATASET 1 .....	44
4.2.2 DATASET 2 .....	47
4.2.3 DATASET 3 .....	51
BAB V KESIMPULAN.....	55
DAFTAR PUSTAKA .....	56
LAMPIRAN.....	62

## DAFTAR GAMBAR

<b>Gambar 3.1</b> Arsitektur Sistem .....	21
<b>Gambar 3.2</b> Dataset 1 .....	24
<b>Gambar 3.3</b> Dataset 2 .....	25
<b>Gambar 3.4</b> Dataset 3 .....	25
<b>Gambar 4.1</b> Confusion Matriks dataset 1 .....	34
<b>Gambar 4.2</b> Confusion Matriks dataset 2 .....	38
<b>Gambar 4.3</b> Confusion Matriks dataset 3 .....	42



## DAFTAR TABEL

<b>Tabel 2.1</b> Interpretasi Nilai Fairness Berdasarkan TPR Ratio .....	16
<b>Tabel 2.2</b> Interpretasi Nilai Fairness Berdasarkan ACC Ratio.....	17
<b>Tabel 2.3</b> Interpretasi Nilai Fairness Berdasarkan PPV Ratio .....	18
<b>Tabel 2.4</b> Interpretasi Nilai Fairness Berdasarkan FPR Ratio.....	19
<b>Tabel 2.5</b> Interpretasi Nilai Fairness Berdasarkan STP Ratio.....	20
<b>Tabel 3.1</b> Atribut Dataset "CDC Diabetes Health Indicators" .....	22
<b>Tabel 4.1</b> Evaluasi model dataset 1 .....	34
<b>Tabel 4.2</b> Evaluasi model dataset 2 .....	37
<b>Tabel 4.3</b> Evaluasi model dataset 3 .....	41
<b>Tabel 4.4</b> Matriks Deteksi Bias Dataset 1 Decision Tree .....	45
<b>Tabel 4.5</b> Matriks Deteksi Bias Dataset 1 Random Forest.....	45
<b>Tabel 4.6</b> Matriks Deteksi Bias Dataset 1 Logistic Regression .....	46
<b>Tabel 4.7</b> Matriks Deteksi Bias Dataset 1 Naive Bayes.....	47
<b>Tabel 4.8</b> Matriks Deteksi Bias Dataset 2 Decision Tree.....	48
<b>Tabel 4.9</b> Matriks Deteksi Bias Dataset 2 Random Forest.....	48
<b>Tabel 4.10</b> Matriks Deteksi Bias Dataset 2 Logistic Regression .....	49
<b>Tabel 4.11</b> Matriks Deteksi Bias Dataset 2 Naive Bayes.....	50
<b>Tabel 4.12</b> Matriks Deteksi Bias Dataset 3 Decision Tree.....	51
<b>Tabel 4.13</b> Matriks Deteksi Bias Dataset 3 Random Forest.....	52
<b>Tabel 4.14</b> Matriks Deteksi Bias Dataset 3 Logistic Regression .....	52
<b>Tabel 4.15</b> Matriks Deteksi Bias Dataset 3 Naive Bayes.....	53

## DAFTAR LAMPIRAN

Lampiran 1. Contoh Perhitungan Matrix Bias .....	62
--	----



## DAFTAR PUSTAKA

- [1] Wang, Z., Gao, C., Xiao, C., & Sun, J. (2024). *MediTab: Scaling Medical Tabular Data Predictors via Data Consolidation, Enrichment, and Refinement*. 6062–6070. <https://doi.org/10.24963/ijcai.2024/670>
- [2] Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., Shaw, J. E., Bright, D., & Williams, R. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Research and Clinical Practice*, 157, 107843. <https://doi.org/10.1016/j.diabres.2019.107843>
- [3] Wardani, N. W. S. (2022). Pemberdayaan Masyarakat Dalam Pencegahan Penyakit Ginjal Pasien Diabetes Dan Hipertensi Di Rsud Sanjiwani Gianyar. *WICAKSANA: Jurnal Lingkungan Dan Pembangunan*, 6(1), 9–16. <https://doi.org/10.22225/wicaksana.6.1.2022.9-16>
- [4] Wijaya, I. P. A., Widarma Atmaja, I. K., & Ayu Septia Dewi, N. K. (2023). Pendidikan Kesehatan Gerakan Masa Depan Sehat Dengan Mencegah Dan Kendalikan Si Manis Berujung Kronis (Diabetes Melitus) Sejak Dini Di Banjar Umahanyar, Abiansemal, Badung. *Jurnal Pengabdian Masyarakat Bina Usada (ABDINUSA)*, 1(1), 20–26. <https://doi.org/10.36474/abdinusa.v1i1.306>
- [5] Yameny, A. A. (2024). Diabetes Mellitus Overview 2024. *Journal of Bioscience and Applied Research*, 10(3), 641–645. <https://doi.org/10.21608/jbaar.2024.382794>
- [6] The Lancet. (2023). Diabetes: a defining disease of the 21st century. *The Lancet*, 401(10394), 2087. [https://doi.org/10.1016/S0140-6736\(23\)01296-5](https://doi.org/10.1016/S0140-6736(23)01296-5)
- [7] Syafaah, L., Basuki, S., Sumadi, F. D. S., Faruq, A., & Purnomo, M. H. (2020). Diabetes prediction based on discrete and continuous mean amplitude of glycemic excursions using machine learning. *Bulletin of Electrical*

*Engineering and Informatics*, 9(6), 2619–2629.  
<https://doi.org/10.11591/eei.v9i6.2387>

- [8] Cesarelli, G., Ponsiglione, A. M., Sansone, M., Amato, F., Donisi, L., & Ricciardi, C. (2024). Machine Learning for Biomedical Applications. *Bioengineering*, 11(8), 1–6. <https://doi.org/10.3390/bioengineering11080790>
- [9] Iparraguirre-Villanueva, O., Espinola-Linares, K., Flores Castañeda, R. O., & Cabanillas-Carbonell, M. (2023). Application of Machine Learning Models for Early Detection and Accurate Classification of Type 2 Diabetes. *Diagnostics*, 13(14). <https://doi.org/10.3390/diagnostics13142383>
- [10] Jain, V., & Tiwari, S. K. (2024). Overview: Machine Learning. *Machine Learning An Art of Computer Thinking*, 130–144. <https://doi.org/10.58532/nbennurch183>
- [11] Aissaoui Ferhi, L., Ben Amar, M., Choubani, F., & Bouallegue, R. (2024). Enhancing diagnostic accuracy in symptom-based health checkers: a comprehensive machine learning approach with clinical vignettes and benchmarking. *Frontiers in Artificial Intelligence*, 7(October). <https://doi.org/10.3389/frai.2024.1397388>
- [12] Raza, S., Pour, P. O., & Bashir, S. R. (2023). Fairness in Machine Learning Meets with Equity in Healthcare. *Proceedings of the Inaugural 2023 Summer Symposium Series 2023*, 149–153. <https://doi.org/10.1609/aaaiss.v1i1.27493>
- [13] Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). Bias and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. *Big Data and Cognitive Computing*, 7(1), 1–31. <https://doi.org/10.3390/bdcc7010015>
- [14] Baniecki, H., Kretowicz, W., Piatyszek, P., Wisniewski, J., & Biecek, P. (2021). dalex: Responsible machine learning with interactive explainability and fairness in python. *Journal of Machine Learning Research*, 22, 1–7.

- [15] Tuppad, A., & Patil, S. D. (2023). Data Pre-processing Issues in Medical Data Classification. *2023 International Conference on Network, Multimedia and Information Technology, NMITCON 2023*, 44, 1079–1084. <https://doi.org/10.1109/NMITCON58196.2023.10275855>
- [16] Prof. Arati K Kale, & Dr. Dev Ras Pandey. (2024). Data Pre-Processing Technique for Enhancing Healthcare Data Quality Using Artificial Intelligence. *International Journal of Scientific Research in Science and Technology*, 299–309. <https://doi.org/10.32628/ijrst52411130>
- [17] Fan, C., Chen, M., Wang, X., Wang, J., & Huang, B. (2021). A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. *Frontiers in Energy Research*, 9(March), 1–17. <https://doi.org/10.3389/fenrg.2021.652801>
- [18] Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/10.1016/j.gltp.2022.04.020>
- [19] Afsaneh, E., Sharifdini, A., Ghazzaghi, H., & Ghobadi, M. Z. (2022). Recent applications of machine learning and deep learning models in the prediction, diagnosis, and management of diabetes: a comprehensive review. *Diabetology and Metabolic Syndrome*, 14(1). <https://doi.org/10.1186/s13098-022-00969-9>
- [20] Mahesh, B. (2020). Machine Learning Algorithms - A Review. *International Journal of Science and Research (IJSR)*, 9(1), 381–386. <https://doi.org/10.21275/art20203995>
- [21] Lehr, J., Philipps, J., Nguyen Hoang, V., von Wrangel, D., & Krüger, J. (2021). Supervised learning vs. unsupervised learning: A comparison for optical inspection applications in quality control. *Proceedings of the International Conference of DAAAM Baltic , 2021-April*. <https://doi.org/10.1088/1757-899X/1140/1/012049>
- [22] Madhuri, M., Samyama Gunjal, G. H., & Kamalapurkar, S. (2020). Air pollution prediction using machine learning supervised learning approach. *International Journal of Scientific and Technology Research*, 9(4), 118–123.

- [23] Bahtiar, S. A. H. (2023). Perbandingan Naïve Bayes dan Logistic Regression dalam Sentiment Analysis pada Review Marketplace menggunakan Rating-Based Labeling. *Universitas Islam Indonesia*.  
dspace.uui.ac.id/123456789/46497
- [24] Sitinjak, J. R., Ichsan, H. H. M., & Setiawan, E. (2023). Penerapan Metode Naive Bayes dalam Sistem Pendeteksi Kualitas Tanah pada Tanaman Kedelai. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 7(5), 2617–2622. <http://j-ptiik.ub.ac.id>
- [25] Ramadhon, R. N., Ogi, A., Agung, A. P., Putra, R., Febrihartina, S. S., & Firdaus, U. (2024). Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank. *Karimah Tauhid*, 3(2), 1860–1874. <https://doi.org/10.30997/karimahtauhid.v3i2.11952>
- [26] Mustofa, Y. A. (2024). Pendekatan Ensemble pada Analisis Sentimen Ulasan Aplikasi Google Play Store Ensemble Approach to Sentiment Analysis of Google Play Store App Reviews. 6, 181–188.
- [27] Aditya, M. F., Pramuntadi, A., Wijaya, D. P., & Wicaksono, Y. (2024). Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(3), 1104–1110. <https://doi.org/10.57152/malcom.v4i3.1284>
- [28] Nurussakinah, N., & Faisal, M. (2023). Klasifikasi Penyakit Diabetes Menggunakan Algoritma Decision Tree. *Jurnal Informatika*, 10(2), 143–149. <https://doi.org/10.31294/inf.v10i2.15989>
- [29] Iskandar, R. F. N., Gutama, D. H., Wijaya, D. P., & Danianti, D. (2024). Klasifikasi Menggunakan Metode Random Forest untuk Awal Deteksi Diabetes Melitus Tipe 2. *Jurnal Teknik Industri Terintegrasi*, 7(3), 1620–1626. <https://doi.org/10.31004/jutin.v7i3.26916>
- [30] Noviyanti, C. N., & Alamsyah, A. (2024). Early Detection of Diabetes Using Random Forest Algorithm. *Journal of Information System Exploration and Research*, 2(1), 41–48. <https://doi.org/10.52465/joiser.v2i1.245>

- [31] Babatunde, R. S., Babatunde, A. N., Balogun, B. F., AbdulRahman, T. A., Umar, E., Ajiboye, R. A., Mohammed, S. B., Oke, A. A., & Obiwusi, K. Y. (2024). A Logistic Regression-Based Technique for Predicting Type II Diabetes. *Advances in Multidisciplinary & Scientific Research Journal Publications*, 4(1), 1–14. <https://doi.org/10.22624/aims/fcsij/2024/p1>
- [32] Al-Rimmawi, H. (2024). Prediction of Type 2 Diabetes using logistic regression techniques. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 15(1), 1–13. <https://doi.org/10.61841/turcomat.v15i1.13875>
- [33] Wickramasinghe, I., & Kalutarage, H. (2021). Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Computing*, 25(3), 2277–2293. <https://doi.org/10.1007/s00500-020-05297-6>
- [34] Setya Pratama, D. P. (2024). *Deteksi Bias dalam Model Machine Learning untuk Prediksi Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual Learning Environment Laporan Tugas Akhir Deteksi Bias dalam Model Machine Learning untuk Prediksi Kelulusan Mahasiswa Berdasarkan Aktivitas Virtual.*
- [35] Wiśniewski, J., & Biecek, P. (2022). fairmodels: a Flexible Tool for Bias Detection, Visualization, and Mitigation in Binary Classification Models. *R Journal*, 14(1), 227–243. <https://doi.org/10.32614/RJ-2022-019>
- [36] Anil Kumar Prajapati, Umesh Kumar Singh, Rekha Singh, & Shukla, A. (2024). Study and analysis of feature selection problems and impact of bias in machine learning disease prediction models. *International Journal of Engineering & Technology*, 13(2), 182–188. <https://doi.org/10.14419/hn1n3h15>
- [37] Wan, M., Zha, D., Liu, N., & Zou, N. (2023). In-Processing Modeling Techniques for Machine Learning Fairness: A Survey. *ACM Transactions on Knowledge Discovery from Data*, 17(3). <https://doi.org/10.1145/3551390>
- [38] Liu, H. (2022). Trustworthy machine learning: Fairness and robustness. *WSDM 2022 - Proceedings of the 15th ACM International Conference on Web*

*Search and Data Mining*, 1553–1554.  
<https://doi.org/10.1145/3488560.3502211>

- [39] Theocharopoulos, P. C., Bersimis, S., Georgakopoulos, S. V., Karaminas, A., Tasoulis, S. K., & Plagianakos, V. P. (2024). Developing predictive precision medicine models by exploiting real-world data using machine learning methods. *Journal of Applied Statistics*.  
<https://doi.org/10.1080/02664763.2024.2315451>



### FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Muhamad Minan Nur Mu'alim

NIM : 202110370311197

Judul TA : PENERAPAN MACHINE LEARNING FAIRNESS PADA  
DATA TABULAR : STUDI KASUS TERHADAP DATA MEDIS

#### Hasil Cek Plagiarisme dengan Turnitin

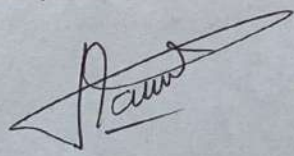
No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	5 %
2.	Bab 2 – Daftar Pustaka	25 %	0 %
3.	Bab 3 – Analisis dan Perancangan	25 %	10 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	0 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	0 %
6.	Makalah Tugas Akhir	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)

  
(..... Agus Muisono .....)