

Klasifikasi Deteksi Penyakit Diabetes Menggunakan Metode Extreme Gradient Boosting (XGBoost) dengan Hyperparameter Tuning

Laporan Tugas Akhir

Diajukan Untuk Memenuhi

Persyaratan Guna Meraih Gelar Sarjana

Informatika Universitas Muhammadiyah Malang



Devi Aprilya Dinanthi

202010370311297

Bidang Minat

Data Science

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH MALANG

2024

LEMBAR PERSETUJUAN

Klasifikasi Deteksi Penyakit Diabetes Menggunakan Metode Extreme Gradient
Boosting (XGBoost) dengan Hyperparameter Tuning

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1

Informatika Universitas Muhammadiyah Malang

Menyetujui,

Malang, 16 Juli 2024

Dosen Pembimbing 1

Dosen Pembimbing 2



Christian Sri Kusuma Aditya S.Kom.,

M.Kom

NIP. 180327021991PNS.

Didih Rizki Chandranegara S.kom.,

M.Kom

NIP. 180302101992PNS.

LEMBAR PENGESAHAN
Klasifikasi Deteksi Penyakit Diabetes Menggunakan Metode
Extreme Gradient Boosting (XGBoost) dengan Hyperparameter
Tuning

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
InformatikaUniversitas Muhammadiyah Malang

Disusun Oleh :
Devi Aprilya Dinanthi
202010370311297

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

Menyetujui,

Dosen Penguji 1



Dosen Penguji 2



Hardianto Wibowo S.Kom, MT.

NIP. 10816120592PNS.

Ir. Galih Wasis Wicaksono S.kom.

M.Cs.

NIP. 10814100541PNS.



Mengetahui,
Ketua Jurusan Informatika
Ir. Galih Wasis Wicaksono S.kom. M.Cs.
NIP. 10814100541PNS.

LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Devi Aprilya Dinanthi

NIM : 202010370311297

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul "**Klasifikasi Deteksi Penyakit Diabetes Menggunakan Metode Extreme Gradient Boosting (XGBoost) dengan Hyperparameter Tuning**" 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

Malang, 16 Juli 2024
Yang Membuat Pernyataan



Christian Sri Kusuma Aditya S.Kom.,
M.Kom

Devi Aprilya Dinanthi

ABSTRAK

Diabetes adalah kondisi serius yang dapat menyebabkan komplikasi fatal dan kematian akibat gangguan metabolisme yang disebabkan oleh kekurangan produksi insulin dalam tubuh. Penelitian ini bertujuan untuk menemukan performa klasifikasi terbaik pada dataset diabetes menggunakan metode Extreme Gradient Boosting (XGBoost). Dataset yang digunakan memiliki 768 baris dan 9 kolom, dengan nilai target 0 untuk pasien tidak diabetes dan 1 untuk pasien diabetes. Dalam penelitian ini, resampling diterapkan untuk mengatasi ketidakseimbangan data menggunakan SMOTE, dan optimasi hyperparameter dilakukan dengan GridSearchCV dan RandomSearchCV. Evaluasi model dilakukan menggunakan confusion matrix seperti akurasi, presisi, recall, dan F1-score. Hasil pengujian menunjukkan bahwa penggunaan GridSearchCV dan RandomSearchCV untuk tuning hyperparameter memberikan hasil yang baik. Penerapan resampling data juga berhasil meningkatkan performa model secara keseluruhan, terutama pada metode XGBoost yang dioptimalkan menggunakan GridSearchCV, yang mencapai akurasi tertinggi sebesar 85%, sementara XGBoost dengan optimasi RandomSearchCV menunjukkan akurasi sebesar 83%.

Kata Kunci : Diabetes, XGBoost, SMOTE, Hyperparameter Tuning, GridSearchCV, RandomSearchCV

ABSTRACT

Diabetes is a serious condition that can cause fatal complications and death due to metabolic disorders caused by insufficient insulin production in the body. This study aims to find the best classification performance on diabetes dataset using Extreme Gradient Boosting (XGBoost) method. The dataset used has 768 rows and 9 columns, with a target value of 0 for non-diabetic patients and 1 for diabetic patients. In this study, resampling was applied to overcome data imbalance using SMOTE, and hyperparameter optimization was performed with GridSearchCV and RandomSearchCV. Model evaluation was performed using confusion matrix as well as metrics such as accuracy, precision, recall, and F1-score. The test results show that the use of GridSearchCV and RandomSearchCV for hyperparameter tuning provides good results. The application of data resampling also managed to improve the overall performance of the model, especially in the XGBoost method optimized using GridSearchCV, which achieved the highest accuracy of 85%, while XGBoost with RandomSearchCV optimization showed an accuracy of 83%.

Keywords : Diabetes, XGBoost, SMOTE, Hyperparameter Tuning, GridSearchCV, RandomSearchCV

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "**Klasifikasi Deteksi Penyakit Diabetes Menggunakan Metode Extreme Gradient Boosting (XGBoost) dengan Hyperparameter Tuning**" yang disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana di Fakultas Teknik, Program Studi Informatika.

Penulis menyadari bahwa penelitian ini masih jauh dari kata sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun untuk penyempurnaan tugas akhir ini di masa yang akan datang. Penelitian ini tentunya tidak lepas dari bimbingan, masukan serta arahan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Bapak/Ibu Dekan Fakultas Teknik, Universitas Muhammadiyah Malang.
2. Bapak Christian Sri Kusuma Aditya S.Kom., M.Kom selaku dosen pembimbing I dan Bapak Didih Rizki Chandranegara S.kom., M.Kom selaku dosen pembimbing II yang selalu memberikan bimbingan, arahan, nasihat serta masukan untuk penulis dalam proses penggerjaan jurnal hingga laporan tugas akhir sehingga dapat terselesaikan dengan baik.
3. Bapak/Ibu Ketua Jurusan Informatika yang telah memberikan kemudahan dan kelancaran semasa studi.
4. Bapak/Ibu Dosen Jurusan Informatika, yang telah memberikan ilmu dan pengetahuan selama penulis menempuh Pendidikan.
5. Kedua orang tua dan keluarga tercinta yang telah memberikan dukungan moral, materi, motivasi dan doa yang selalu di berikan.
6. Sahabat-sahabat seperjuangan penulis, atas dukungan, motivasi, hiburan, kebersamaan, kerja sama selama ini.

Malang, 19 Juni 2024



Devi Aprilya Dinanthi

LEMBAR PERSEMBAHAN

Tugas akhir ini saya persembahkan dengan penuh rasa cinta dan rasa syukur kepada :

1. Kedua Orang Tua Tercinta Bapak Zainal Arifin dan Ibu Amnah, yang telah membesarakan penulis hingga saat ini. Terima kasih banyak atas dukungan moral dan materi yang selalu di berikan, serta doa dan motivasi yang tak henti-hentinya mengalir kepada penulis, sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik.
2. Elisa Ramadanti, penulis ucapkan terima kasih karena sudah membersamai selama penggerjaan jurnal.
3. Adikku Tercinta Diva Febrina Dinanthi, yang selalu mendukung dan memberikan motivasi penulis selama menjalani perkuliahan dan menyelesaikan tugas akhir ini
4. Kakaku tersayang Thursina Megah, yang selalu memberikan nasihat, dukungan, dan dorongan selama penulis menjalani perkuliahan dan mengerjakan tugas akhir ini. Terima kasih telah selalu menemani dalam penggerjaan tugas akhir ini.
5. Teman-teman dan Sahabat-sahabat Tercinta Yayan, Fila, Novi, Nisa, Dini, Elan, dan Viona, yang selalu memberikan semangat, bantuan, motivasi, dan kerja sama yang baik selama saya mengerjakan tugas akhir ini. Terima kasih atas persahabatan dan dukungan kalian yang tiada henti.
6. Partner-partner yang tidak dapat disebutkan satu persatu, terima kasih kepada semua partner yang telah berkontribusi dan memberikan bantuan selama proses penggerjaan tugas akhir ini. Dukungan dan kerjasama kalian sangat berarti bagi saya. Semoga kebaikan kalian mendapatkan balasan yang setimpal.
7. Diriku sendiri, terima kasih atas ketekunan, kesabaran, dan kerja keras yang telah dilakukan. Setiap usaha, pengorbanan, dan dedikasi telah membawa hasil yang membanggakan. Semoga pencapaian ini menjadi langkah awal menuju kesuksesan yang lebih besar di masa depan dan semoga kedepannya bisa berkembang lebih baik jauh lagi.

Akhir kata, semoga tugas akhir ini dapat memberikan manfaat dan menjadi langkah awal untuk meraih kesuksesan di masa depan. Semoga Allah SWT selalu memberkati dan melimpahkan rahmat-Nya kepada kita semua.

Malang, 19 Juni 2024

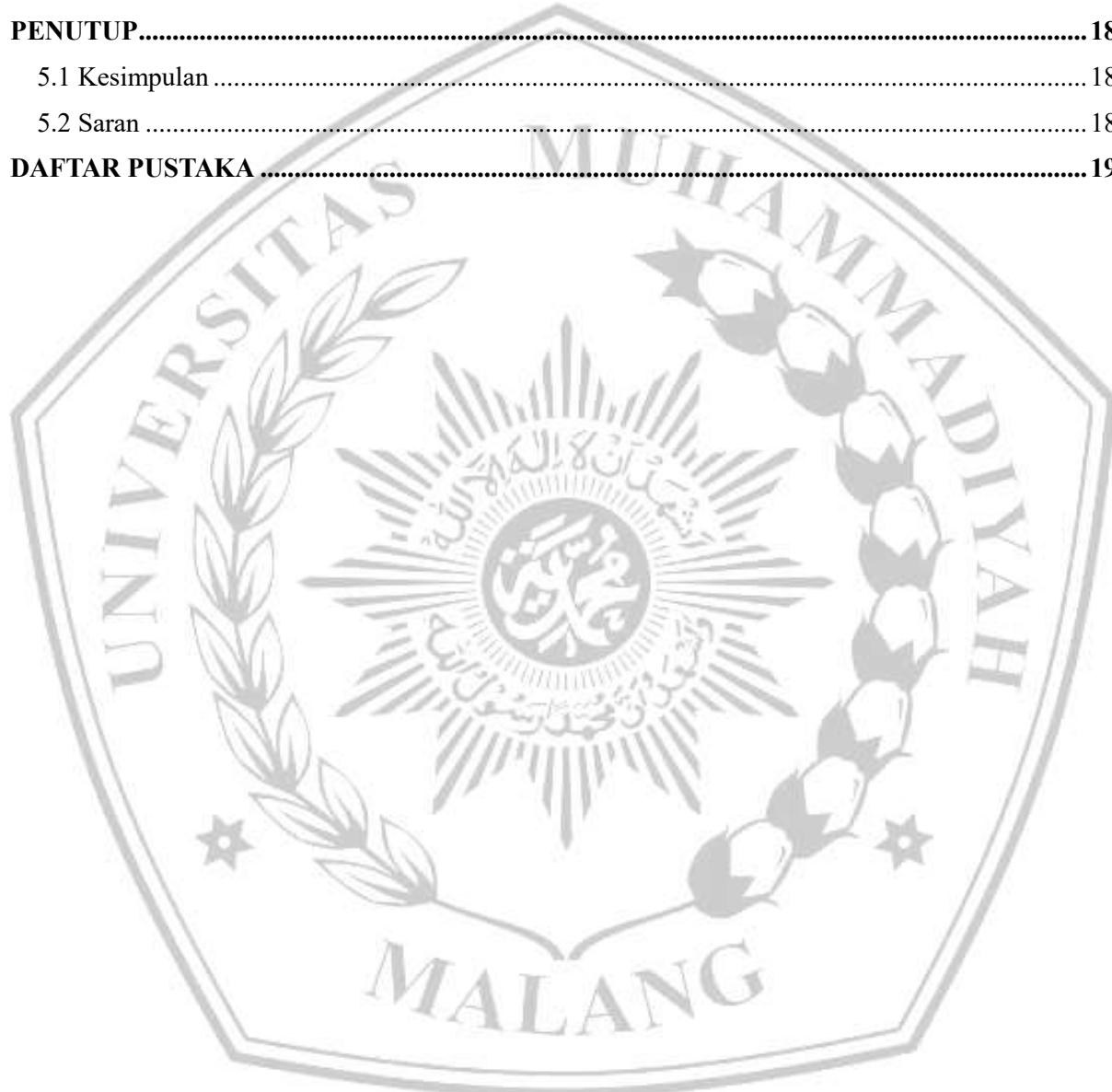


Devi Aprilya Dinanthi

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
LEMBAR PERNYATAAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	v
LEMBAR PENGESAHAN.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
BAB 1	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
BAB II	4
TINJAUAN PUSTAKA	4
2.1 Penelitian Terdahulu	4
2.2 Diabetes.....	7
2.3 Hyperparameter Tuning	7
2.3.1 GridSearchCV	7
2.3.2 RandomSearchCV	8
2.4 Extreme Gradient Boosting (XGBoost).....	8
BAB III.....	9
METODOLOGI PENELITIAN	9
3.1 Rancangan Penelitian.....	9
3.2 Dataset.....	9
3.3 Preprocessing	10
3.4 Data Cleaning.....	10
3.5 Data Normalisasi.....	11
3.6 Resampling	11
3.7 Hyperparameter Tuning	12
3.8 Metode Extreme Gradient Boosting (XGBoost).....	12
3.9 Evaluasi.....	12
BAB IV	14

HASIL DAN PEMBAHASAN	14
4.1 Implementasi Metode GridSearchCV	14
4.1.1 Hasil Klasifikasi Metode GridSearchCV	14
4.2 Implementasi Metode RandomSearchCV	15
4.2.1 Hasil Klasifikasi Metode RandomSearchCV.....	15
4.3 Implementasi Metode Resampling.....	16
BAB V	18
PENUTUP.....	18
5.1 Kesimpulan	18
5.2 Saran	18
DAFTAR PUSTAKA	19



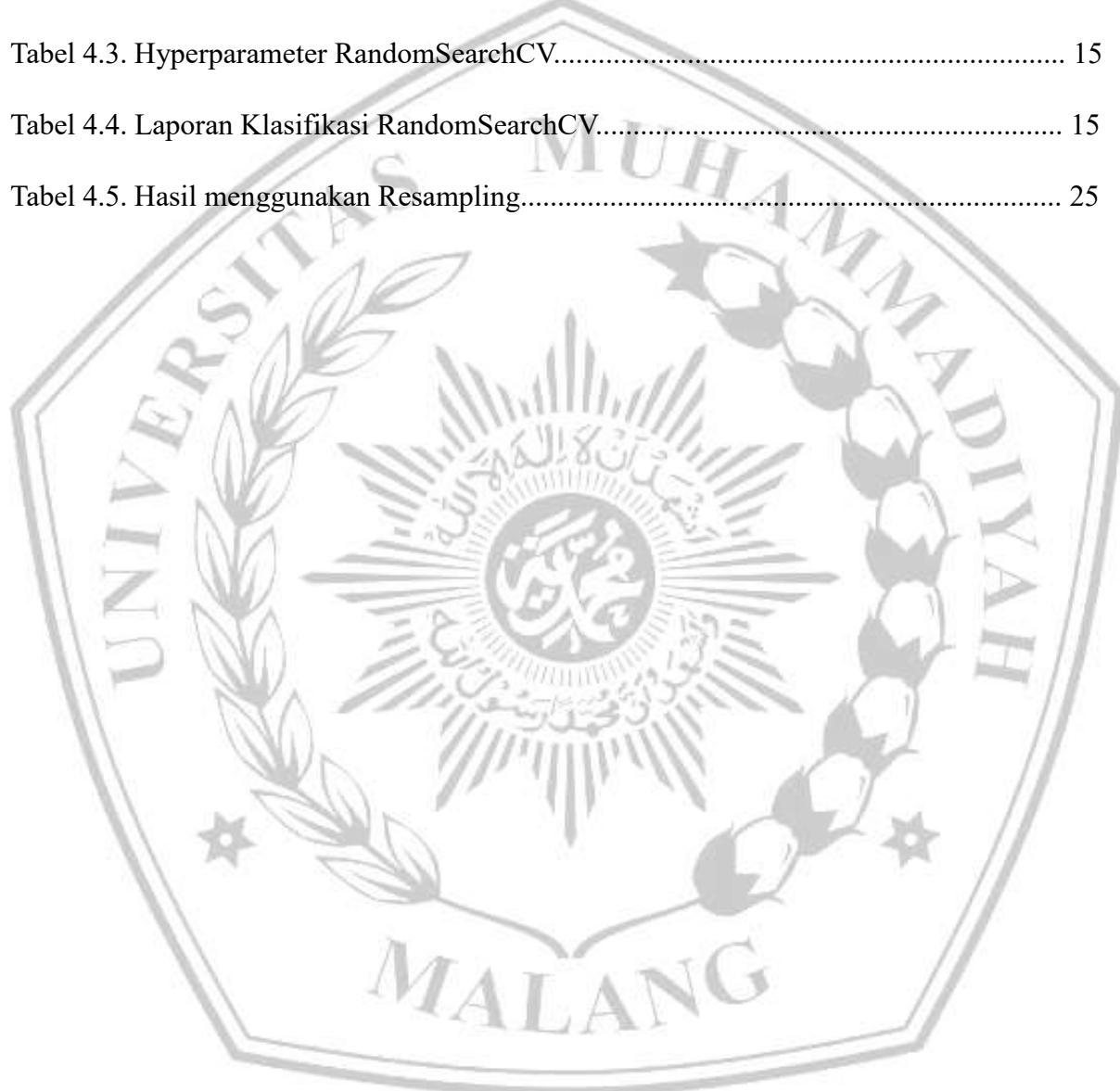
DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian	9
Gambar 3.2. Sebelum Resampling.....	10
Gambar 3.3. Setelah Resampling.....	19
Gambar 4.1. Perbedaan hasil XGBoost.....	20



DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu.....	4
Tabel 2.2. Rincian Dataset.....	10
Tabel 4.1. Hyperparameter GridSearchCV	14
Tabel 4.2. Laporan Klasifikasi dengan GridSearch CV.....	14
Tabel 4.3. Hyperparameter RandomSearchCV.....	15
Tabel 4.4. Laporan Klasifikasi RandomSearchCV.....	15
Tabel 4.5. Hasil menggunakan Resampling.....	25



DAFTAR PUSTAKA

- [1] Y. Mukhtar, A. Galalain, and U. Yunusa, “a Modern Overview on Diabetes Mellitus: a Chronic Endocrine Disorder,” *Eur. J. Biol.*, vol. 5, no. 2, pp. 1–14, 2020, doi: 10.47672/ejb.409.
- [2] M. R. Afandi and F. R. Marpaung, “Correlation Between Apoprotein B/Apoprotein a-I Ratio With Homa Ir Value (Homeostatic Model Assesment Insulin Resistance) in Type 2 Diabetes Mellitus,” *J. Vocat. Heal. Stud.*, vol. 3, no. 2, p. 78, 2019, doi: 10.20473/jvhs.v3.i2.2019.78-82.
- [3] L. Ismail, H. Materwala, and J. Al Kaabi, “Association of risk factors with type 2 diabetes: A systematic review,” *Comput. Struct. Biotechnol. J.*, vol. 19, pp. 1759–1785, 2021, doi: 10.1016/j.csbj.2021.03.003.
- [4] M. J. Hossain, M. Al-Mamun, and M. R. Islam, “Diabetes mellitus, the fastest growing global public health concern: Early detection should be focused,” *Heal. Sci. Reports*, vol. 7, no. 3, pp. 5–9, 2024, doi: 10.1002/hsr2.2004.
- [5] S. Webber, *International Diabetes Federation*, vol. 102, no. 2. 2013. doi: 10.1016/j.diabres.2013.10.013.
- [6] B. CURCHOD and J. P. DAEPSEN, “Complications of diabetes mellitus,” *Praxis (Bern. 1994)*., vol. 48, no. 26, pp. 602–603, 1959.
- [7] Erlin, Y. N. Marlism, Junadhi, L. Suryati, and N. Agustina, “Early Detection of Diabetes Using Machine Learning with Logistic Regression Algorithm,” *J. Nas. Tek. Elektro dan Teknol. Inf.*, vol. 11, no. 2, pp. 88–96, 2022.
- [8] K. M. Aamir, L. Sarfraz, M. Ramzan, M. Bilal, J. Shafi, and M. Attique, “A fuzzy rule-based system for classification of diabetes,” *Sensors*, vol. 21, no. 23, 2021, doi: 10.3390/s21238095.
- [9] H. Hairani and D. Priyanto, “A New Approach of Hybrid Sampling SMOTE and ENN to the Accuracy of Machine Learning Methods on Unbalanced Diabetes Disease Data,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 585–590, 2023, doi: 10.14569/IJACSA.2023.0140864.
- [10] Y. Sitinjak, M. Nababan, and M. City, “Liver Disease Classification Analysis,” vol. 7, no. 1, pp. 132–141, 2023.

- [11] F. Kanwal, M. K. Abid, M. S. Maqbool, D. N. Aslam, and M. Fuzail, “Optimized Classification of Cardiovascular Disease Using Machine Learning Paradigms,” *VFAST Trans. Softw. Eng.*, vol. 11, no. 2, pp. 140–148, 2023, doi: 10.21015/vtse.v11i2.1527.
- [12] Rahmanul Hoque, Suman Das, Mahmudul Hoque, and Mahmudul Hoque, “Breast Cancer Classification using XGBoost,” *World J. Adv. Res. Rev.*, vol. 21, no. 2, pp. 1985–1994, 2024, doi: 10.30574/wjarr.2024.21.2.0625.
- [13] L. Liao, H. Li, W. Shang, and L. Ma, “An Empirical Study of the Impact of Hyperparameter Tuning and Model Optimization on the Performance Properties of Deep Neural Networks,” *ACM Trans. Softw. Eng. Methodol.*, vol. 31, no. 3, pp. 1–39, 2022, doi: 10.1145/3506695.
- [14] Lestari, Zulkarnain, and S. A. Sijid, “Diabetes Melitus: Review Etiologi, Patofisiologi, Gejala, Penyebab, Cara Pemeriksaan, Cara Pengobatan dan Cara Pencegahan,” *UIN Alauddin Makassar*, no. November, pp. 237–241, 2021, [Online]. Available: <http://journal.uin-alauddin.ac.id/index.php/psb>
- [15] L. Silalahi, “Hubungan Pengetahuan dan Tindakan Pencegahan Diabetes Mellitus Tipe 2,” *J. PROMKES*, vol. 7, no. 2, p. 223, 2019, doi: 10.20473/jpk.v7.i2.2019.223-232.
- [16] A. Khadilkar and C. Oza, “Glycaemic Control in Youth and Young Adults: Challenges and Solutions,” *Diabetes, Metab. Syndr. Obes.*, vol. 15, pp. 121–129, 2022, doi: 10.2147/DMSO.S304347.
- [17] Y. Zheng, S. H. Ley, and F. B. Hu, “Global aetiology and epidemiology of type 2 diabetes mellitus and its complications,” *Nat. Rev. Endocrinol.*, vol. 14, no. 2, pp. 88–98, 2018, doi: 10.1038/nrendo.2017.151.
- [18] W. Nugraha and A. Sasongko, “Hyperparameter Tuning pada Algoritma Klasifikasi dengan Grid Search,” *Sist. J. Sist. Inf.*, vol. 11, no. 2, pp. 391–401, 2022, [Online]. Available: <http://sistemasasi.ftik.unisi.ac.id>
- [19] M. Fajri and A. Primajaya, “Komparasi Teknik Hyperparameter Optimization pada SVM untuk Permasalahan Klasifikasi dengan Menggunakan Grid Search dan Random Search,” *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 14–19, 2023, doi: 10.30871/jaic.v7i1.5004.
- [20] R. Valarmathi and T. Sheela, “Heart disease prediction using hyper parameter

optimization (HPO) tuning,” *Biomed. Signal Process. Control*, vol. 70, no. March, p. 103033, 2021, doi: 10.1016/j.bspc.2021.103033.

- [21] W. A. Kusuma and P. Korespondensi, “PREDIKSI INTERAKSI DRUG TARGET PADA GEN KANKER MENGGUNAKAN METODE LASSO-XGBOOST DRUG TARGET INTERACTION PREDICITION ON CANCER GENE USING LASSO-XGBOOST METHOD,” vol. 10, no. 3, pp. 531–542, 2023, doi: 10.25126/jtiik.2023106603.
- [22] M. Guo, Z. Yuan, B. Janson, Y. Peng, Y. Yang, and W. Wang, “Older pedestrian traffic crashes severity analysis based on an emerging machine learning xgboost,” *Sustain.*, vol. 13, no. 2, pp. 1–26, Jan. 2021, doi: 10.3390/su13020926.
- [23] M. R. Kurniawanda and F. A. T. Tobing, “Analysis Sentiment Cyberbullying In Instagram Comments with XGBoost Method,” *IJNMT (International J. New Media Technol.)*, vol. 9, no. 1, pp. 28–34, 2022, doi: 10.31937/ijnmt.v9i1.2670.
- [24] F. Putra, H. F. Tahiyat, and R. M. Ihsan, “Application of K-Nearest Neighbor Algorithm Using Wrapper as Preprocessing for Determination of Human Weight Information Penerapan Algoritma K-Nearest Neighbor Menggunakan Wrapper Sebagai Preprocessing untuk Penentuan Keterangan Berat Badan Manusia,” vol. 4, no. January, pp. 273–281, 2024.
- [25] F. Neutatz, B. Chen, Y. Alkhatib, J. Ye, and Z. Abedjan, “Data Cleaning and AutoML: Would an Optimizer Choose to Clean?,” *Datenbank-Spektrum*, vol. 22, no. 2, pp. 121–130, 2022, doi: 10.1007/s13222-022-00413-2.
- [26] F. N. S. Inggih Permana, “The Effect of Data Normalization on the Performance of the Classification Results of the Backpropagation Algorithm,” *IJRSE Indones. J. Inform. Res. Softw. Eng.*, vol. 2, no. 1, pp. 67–72, 2022.
- [27] D. Singh and B. Singh, “Investigating the impact of data normalization on classification performance,” *Appl. Soft Comput.*, vol. 97, no. xxxx, p. 105524, 2020, doi: 10.1016/j.asoc.2019.105524.
- [28] R. Ghorbani and R. Ghousi, “Comparing Different Resampling Methods in Predicting Students’ Performance Using Machine Learning Techniques,” *IEEE Access*, vol. 8, pp. 67899–67911, 2020, doi: 10.1109/ACCESS.2020.2986809.

- [29] J. Han, K. Shu, and Z. Wang, “Predicting energy use in construction using Extreme Gradient Boosting,” *PeerJ Comput. Sci.*, vol. 9, pp. 1–14, 2023, doi: 10.7717/peerj-cs.1500.
- [30] F. Nurrahman, H. Wijayanto, A. H. Wigena, and N. Nurjanah, “Pre-Processing Data on Multiclass Classification of Anemia and Iron Deficiency With the Xgboost Method,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 17, no. 2, pp. 0767–0774, 2023, doi: 10.30598/barekengvol17iss2pp0767-0774.
- [31] N. T. Tran, T. T. G. Tran, T. A. Nguyen, and M. B. Lam, “A new grid search algorithm based on XGBoost model for load forecasting,” *Bull. Electr. Eng. Informatics*, vol. 12, no. 4, pp. 1857–1866, 2023, doi: 10.11591/eei.v12i4.5016.
- [32] J. H. B, “Risk Level Prediction of Life Insurance Applicant using Machine Learning,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 2213–2220, 2020, doi: 10.30534/ijatcse/2020/199922020.





UNIVERSITAS
MUHAMMADIYAH
MALANG

FAKULTAS TEKNIK

INFORMATIKA

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



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Devi Aprilya Dinanthi

NIM : 202010370311297

Judul TA : Klasifikasi Deteksi Penyakit Diabetes Menggunakan Metode Extreme Gradient Boosting (XGBoost) dengan Hyperparameter Tuning

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	7%
2.	Bab 2 – Daftar Pustaka	25 %	20%
3.	Bab 3 – Analisis dan Perancangan	25 %	23%
4.	Bab 4 – Implementasi dan Pengujian	15 %	12%
5.	Bab 5 – Kesimpulan dan Saran	5 %	3%
6.	Makalah Tugas Akhir	20%	9%

*) Hasil cek plagiarism dilakukan oleh pemeriksa (staff TU)

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

Mengetahui,

Pemeriksa (Staff TU)

(.....)



Kampus I
Jl. Rambutan 1 Malang, Jawa Timur
P: +62 341 591 252 (Hunting)
F: +62 341 469 435

Kampus II
Jl. Rembang-Sukun No.168 Malang, Jawa Timur
P: +62 341 531 149 (Hunting)
F: +62 341 582 060

Kampus III
Jl. Raya Teponmas No.240 Malang, Jawa Timur
P: +62 341 804 219 (Hunting)
F: +62 341 463 435
E: webmaster@umm.ac.id