

**PENERAPAN ALGORITMA *EXTREME GRADIENT BOOSTING*
(XGBOOST) DALAM KLASIFIKASI *CELIAC DISEASE***

SKRIPSI



**UIN SUNAN AMPEL
S U R A B A Y A**

Disusun Oleh
ROUDLOTUL JANNAH ALFIRDAUSY
09030220048

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2024

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : ROUDLOTUL JANNAH ALFIRDAUSY

NIM : 09030220048

Program Studi : Matematika

Angkatan : 2020

Menyatakan bahwa Saya tidak melakukan plagiat dalam penulisan skripsi Saya yang berjudul "PENERAPAN ALGORITMA *EXTREME GRADIENT BOOSTING* (XGBOOST) DALAM KLASIFIKASI *CELIAC DISEASE*". Apabila suatu saat nanti terbukti Saya melakukan tindakan plagiat, maka Saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini Saya buat dengan sebenar-benarnya.

Surabaya, 19 Desember 2023

Yang menyatakan,



ROUDLOTUL JANNAH ALFIRDAUSY
NIM. 09030220048

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : ROUDLOTUL JANNAH ALFIRDAUSY
NIM : 09030220048
Judul Skripsi : PENERAPAN ALGORITMA *EXTREME GRADIENT BOOSTING* (XGBOOST) DALAM KLASIFIKASI *CELIAC DISEASE*

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



Nurissaidah Ulinuha, M.Kom.
NIP. 199011022014032004

Pembimbing II



Wika Dianita Utami, M.Sc.
NIP. 199206102018012003

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya



Yuniar Farida, M.T.
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh

Nama : ROUDLOTUL JANNAH ALFIRDAUSY
NIM : 09030220048
Judul Skripsi : PENERAPAN ALGORITMA *EXTREME GRADIENT BOOSTING* (XGBOOST) DALAM KLASIFIKASI *CELIAC DISEASE*

Telah dipertahankan di depan Tim Penguji
pada tanggal 19 Desember 2023

Mengesahkan,
Tim Penguji

Penguji I



Aris Fanani, M.Kom
NIP. 198701272014031002

Penguji II

Dr. Moh. Hafiyusholeh, M.Si., M.PMat.
NIP. 198002042014031001

Penguji III



Nurissaidah Ulinnuha, M.Kom.
NIP. 199011022014032004

Penguji IV



Wika Dianita Utami, M.Sc.
NIP. 199206102018012003

Mengetahui,

Dekan Fakultas Sains dan Teknologi
UIN Sunan Ampel Surabaya



Agus Hamdani, M.Pd.
NIP. 196507312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpustakaan@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : ROUDLOTUL JANNAH ALFIRDAUSY
NIM : 09030220048
Fakultas/Jurusan : SAINS DAN TEKNOLOGI / MATEMATIKA
E-mail address : firdausy123@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

Skripsi Tesis Desertasi Lain-lain (.....)

yang berjudul :

PENERAPAN ALGORITMA EXTREME GRADIENT BOOSTING
(XGBOOST) DALAM KLASIFIKASI CELIAC DISEASE

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 09 JANUARI 2024

Penulis

(ROUDLOTUL JANNAH A.)

nama terang dan tanda tangan

ABSTRAK

PENERAPAN ALGORITMA *EXTREME GRADIENT BOOSTING* (XGBOOST) DALAM KLASIFIKASI *CELIAC DISEASE*

Celiac Disease (CeD) merupakan gangguan autoimun yang dipicu oleh konsumsi gluten yang berinteraksi dengan sistem imun di usus melalui *Human Leukocyte Antigen* (HLA). Kejadian rata-rata *Celiac Disease* dilaporkan antara 0,5% hingga 1% di seluruh dunia, namun hanya sekitar 30% kasus yang didiagnosis dengan benar. Diagnosis *Celiac Disease* saat ini masih menjadi tantangan bagi para profesional medis. Proses diagnosis yang akurat dan tepat waktu sangat penting untuk menghindari komplikasi yang lebih serius. Namun, diagnosis CeD sering kali memerlukan serangkaian tes yang rumit, termasuk tes darah, biopsi usus halus, dan eliminasi gluten dari diet. XGBoost merupakan teknik pembelajaran mesin yang memanfaatkan *ensemble learning* yang menggunakan *decision tree* sebagai *base learner*. Pada penelitian ini menggunakan metode XGBoost sebagai pembelajaran mesin yang akan digunakan untuk mengklasifikasi *Celiac Disease* dengan akurat dan efisien. Penelitian ini melakukan uji coba pada *max_depth*, *n_estimator*, *gamma* dan *learning rate* menggunakan *5-fold crossvalidation*. *Tuning parameter* yang dilakukan berhasil memberikan peningkatan sebesar 0.45% dari nilai akurasi dengan parameter default XGBoost sebesar 98.19%. Setelah dilakukan *tuning parameter* didapatkan hasil evaluasi terbaik dalam uji coba pada parameter *max_depth* sebesar 3, *n_estimator* sebanyak 100, *gamma* sebesar 0 dan *learning rate* adalah 0.3 dan 0.5. Model tersebut menghasilkan tingkat akurasi sebesar 98.64%, sensitivitas mencapai 98.43% dan spesifisitas sebesar 99.72%.

Kata kunci: Klasifikasi, XGBoost, *Celiac Disease*.

ABSTRACT

APPLICATION OF THE EXTREME GRADIENT BOOSTING (XGBOOST) ALGORITHM IN CELIAC DISEASE CLASSIFICATION

Celiac Disease (CeD) is an autoimmune disorder triggered gluten consumption interacts with the immune system in the intestine through Human Leukocyte Antigen (HLA). The average incidence of Celiac Disease is reported to be between 0.5% to 1% worldwide, but only about 30% of cases are diagnosed correctly. Diagnosis of Celiac Disease is currently still a challenge for medical professionals. An accurate and timely diagnosis process is very important to avoid more serious complications. However, the diagnosis of CeD often requires a complex series of tests, including blood tests, small bowel biopsy, and elimination of gluten from the diet. XGBoost is a machine learning technique that utilizes ensemble learning which uses a decision tree as a base learner. In this study, the XGBoost method is used as machine learning which will be used to classify Celiac Disease accurately and efficiently. This research conducted trials on max depth, n estimator, gamma and learning rate using 5-fold cross validation. The parameter tuning carried out succeeded in providing an increase of 0.45% from the accuracy value with the default XGBoost parameter of 98.19%. After tuning the parameters, the best evaluation results were obtained in the trial with parameters max depth of 3, n_estimator of 100, gamma of 0 and learning rate of 0.3 and 0.5. This model produces an accuracy rate of 98.64%, sensitivity of 98.43% and specificity of 99.72%.

Keywords: Classification, XGBoost, Celiac Disease.

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
KATA PENGANTAR	vi
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
ABSTRAK	xiv
ABSTRACT	xv
I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	8
1.3. Tujuan Penelitian	8
1.4. Manfaat Penelitian	9
1.5. Batasan Masalah	9
1.6. Sistematika Penulisan	9
II TINJAUAN PUSTAKA	11
2.1. <i>Celiac Disease</i>	11
2.2. Klasifikasi	15
2.3. <i>Boosting</i>	16
2.4. <i>Extreme Gradient Boosting</i>	18
2.5. <i>K-Fold Cross Validation</i>	30
2.6. Evaluasi	30
2.7. Integrasi Keislaman	33
III METODE PENELITIAN	37

<http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/>

3.1. Jenis Penelitian	37
3.2. Data Penelitian	37
3.3. Tahapan Penelitian	38
3.3.1. Tahap Analisis Data	38
3.3.2. Tahap Analisis Metode XGBoost	41
3.3.3. Skema Uji Coba Parameter	43
IV HASIL DAN PEMBAHASAN	44
4.1. Deskripsi Data	44
4.2. Pelatihan Model XGBoost	49
4.2.1. Menentukan <i>Initial Prediction</i> dan Residual (<i>Error</i>)	50
4.2.2. Menentukan <i>Similarity Score</i>	52
4.2.3. Penentuan <i>Branch Awal</i>	53
4.2.4. Rekonstruksi Pembentukan Pohon Keputusan XGBoost	61
4.2.5. <i>Prunning</i>	78
4.2.6. Nilai Cover	78
4.2.7. Nilai Output Semua Daun (<i>Leaf Node</i>)	79
4.3. Pengujian Model XGBoost	83
4.4. Uji Coba Parameter XGBoost	86
4.4.1. Klasifikasi XGBoost dengan <i>Default</i> Parameter	87
4.4.2. Klasifikasi XGBoost dengan <i>Tuning</i> Parameter	93
4.4.3. Interpretasi Parameter XGBoost	103
4.5. Integrasi Keilmuan	107
V PENUTUP	110
5.1. Kesimpulan	110
5.2. Saran	111
DAFTAR PUSTAKA	111

DAFTAR TABEL

3.1	Deskripsi Variabel Data <i>Celiac Disease</i>	38
3.2	Deskripsi Parameter Uji Coba XGboost	43
4.1	Statistik Deskriptif Fitur Numerik	44
4.2	<i>Encoding</i> Data Kategorikal	48
4.3	Sampel Acak Data Pelatihan	49
4.4	Nilai Residual (<i>Error</i>) Data Pelatihan	51
4.5	Nilai <i>Similarity</i> dan <i>Gain</i> pada <i>Branch</i> Pertama	60
4.6	Nilai <i>Similarity</i> dan <i>Gain</i> pada Cabang True <i>Max_depth</i> = 2	66
4.7	Nilai <i>Similarity</i> dan <i>Gain</i> pada Cabang False <i>Max_depth</i> = 2	71
4.8	Sampel Acak Data Pengujian	83
4.9	<i>Output Value Testing</i> Model XGBoost	84
4.10	Hasil Prediksi Model XGBoost	86
4.11	Uji Coba Parameter Klasifikasi XGBoost dengan kedalaman pohon = 3	93
4.12	Uji Coba Parameter Klasifikasi XGBoost dengan kedalaman pohon = 6	95
4.13	Uji Coba Parameter Klasifikasi XGBoost dengan kedalaman pohon = 9	96

DAFTAR GAMBAR

2.1 Endoskopi Usus <i>Celiac Disease</i>	13
2.2 Arsitektur Algoritma <i>Boosting</i>	17
2.3 Pohon Keputusan Pembentuk XGBoost	18
2.4 <i>K-Fold Cross Validation</i>	30
2.5 Representasi <i>Confussion Matrix Multiclass</i>	31
3.1 Diagram Alir (<i>Flowchart</i>) Analisis Data <i>Celiac Disease</i> dengan Metode XGBoost	39
3.2 Diagram Alir (<i>Flowchart</i>) <i>Extreme Gradient Boosting</i>	41
4.1 Diagram Sankey Fitur Numerik <i>Celiac Disease</i>	46
4.2 <i>Branch</i> Awal Fitur <i>Gender < 1.5</i>	54
4.3 <i>Branch</i> Awal Fitur <i>Diabetes < 0.5</i>	55
4.4 <i>Branch</i> Awal Fitur <i>Diarrhoea < 1.5</i>	56
4.5 <i>Branch</i> Awal Fitur <i>Diarrhoea < 2.5</i>	57
4.6 <i>Branch</i> Awal Fitur <i>Abdominal < 0.5</i>	59
4.7 <i>Branch</i> Awal Pohon Keputusan XGBoost	60
4.8 <i>Branch</i> Max_depth = 2 Fitur True <i>Gender < 1.5</i>	62
4.9 <i>Branch</i> Max_depth = 2 Fitur True <i>Diabetes < 0.5</i>	63
4.10 <i>Branch</i> Max_depth = 2 Fitur True <i>Diarrhoea < 1.5</i>	64
4.11 <i>Branch</i> Max_depth = 2 Fitur True <i>Diarrhoea < 2.5</i>	64
4.12 <i>Branch</i> Max_depth = 2 Fitur True <i>Abdominal < 0.5</i>	65
4.13 <i>Branch</i> Max_depth = 2 Fitur False <i>Gender < 1.5</i>	67
4.14 <i>Branch</i> Max_depth = 2 Fitur False <i>Diabetes < 0.5</i>	68
4.15 <i>Branch</i> Max_depth = 2 Fitur False <i>Diarrhoea < 1.5</i>	69
4.16 <i>Branch</i> Max_depth = 2 Fitur False <i>Diarrhoea < 2.5</i>	69
4.17 <i>Branch</i> Max_depth = 2 Fitur False <i>Abdominal < 0.5</i>	70
4.18 Konstruksi Pohon Keputusan XGBoost dengan (<i>max_depth</i>) = 2	72

<http://digilib.uinsa.ac.id/> / <http://digilib.uinsa.ac.id/> / <http://digilib.uinsa.ac.id/>

4.19 Pohon dengan Gain Maksimal pada <i>Max Depth =3 Branch</i>	
<i>Gender<1.5 True</i>	73
4.20 Pohon dengan Gain Maksimal pada <i>Max Depth =3 Branch</i>	
<i>Gender<1.5 False</i>	74
4.21 Pohon dengan Gain Maksimal pada <i>Max Depth =3 Branch</i>	
<i>Diarrhoea<2.5 True</i>	75
4.22 Pohon dengan Gain Maksimal pada <i>Max Depth =3 Branch</i>	
<i>Diarrhoea<2.5 False</i>	76
4.23 Pohon Keputusan Pelatihan XGBoost dengan (<i>Max_depth</i>) = 3	77
4.24 Pohon Keputusan XGBoost	82
4.25 Pohon XGBoost pada Klasifikasi <i>Celiac Disease</i> dengan Parameter	
<i>Default</i>	87
4.26 <i>Confussion Matrix</i> Klasifikasi <i>Celiac Disease</i> dengan Parameter	
<i>Default</i> Menggunakan Metode XGBoost	88
4.27 <i>Confussion Matrix</i> Klasifikasi <i>Celiac Disease</i> dengan <i>Tuning</i>	
<i>Parameter</i> Menggunakan Metode XGBoost	98
4.28 Grafik Pengaruh <i>Learning Rate</i> Terhadap Presentase Rata-Rata	
<i>Akurasi</i>	103
4.29 Grafik Pengaruh <i>Gamma</i> Terhadap Presentase Rata-Rata <i>Akurasi</i>	104
4.30 Grafik Pengaruh <i>Max_depth</i> Terhadap Presentase Rata-Rata <i>Akurasi</i>	105
4.31 Grafik Pengaruh <i>N_estimator</i> Terhadap Presentase Rata-Rata <i>Akurasi</i>	106

DAFTAR PUSTAKA

- Alfenda, Y. (2021). Uji suka brownies singkong free gluten. *Jurnal Pesona Hospitality*, 14(2):22–29.
- Amelia, Y. (2023). Perbandingan Metode Machine Learning untuk Mendeteksi Penyakit Jantung. *Idealis : Indonesia Journal Information System*, 6:220–225.
- Bentéjac, C., Csörgő, A., and Martínez-Muñoz, G. (2019). A Comparative Analysis of XGBoost. (November 2019).
- Caratelli, V., Moccia, M., Paggiaro, F. R., Fiore, L., Avitabile, C., Saviano, M., Imbriani, A. L., Dardano, P., De Stefano, L., Moscone, D., Colabufo, N. A., Ghafir El Idrissi, I., Russo, F., Riezzo, G., Giannelli, G., and Arduini, F. (2022). Liquid Biopsy beyond Cancer: A miRNA Detection in Serum with Electrochemical Chip for Non-Invasive Coeliac Disease Diagnosis. *Advanced NanoBiomed Research*, 2(9):2200015.
- Carmona, P., Climent, F., and Momparler, A. (2019). Predicting failure in the U.S. banking sector: An extreme gradient boosting approach. *International Review of Economics and Finance*, 61:304–323.
- Christanto, H., Rahmad, J., Sinurat, S. H., and Ryan, D. (2023). Analisis Perbandingan Decision Tree , Support Vector Machine , dan Xgboost dalam Mengklasifikasi Review Hotel Trip Advisor. *Jurnal Pesona Hospitality*, 9(1):306–319.
- Ciacchi, L., Reid, H. H., and Rossjohn, J. (2022). Structural bases of T cell antigen <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/>

receptor recognition in celiac disease. *Current Opinion in Structural Biology*, 74:102349.

Climent, F., Momparler, A., and Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient Boosting approach. *Journal of Business Research*, 101(June):885–896.

Dani, A. T. R., Ratnasari, V., Ni'matuzzahroh, L., Aviantholib, I. C., Novidianto, R., and Adrianingsih, N. Y. (2022). Analisis Klasifikasi Artist Music Menggunakan Model Regresi Logistik Biner Dan Analisis Diskriminan. *Jambura Journal of Probability and Statistics*, 3(1):1–10.

Demir, S. (2023). An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost , gradient boosting , and XGBoost. *Neural Computing and Applications*, 35(4):3173–3190.

Devan, P. and Khare, N. (2020). An efficient XGBoost–DNN-based classification model for network intrusion detection system. *Neural Computing and Applications*, 32(16):12499–12514.

Diantika, S., Nalatissifa, H., Supriyadi, R., Maulidah, N., and Fauzi, A. (2023). Implementasi Multi-Class Gradient Boosting untuk Mengklasifikasikan Jenis Hewan Pada Kebun Binatang. *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika*, 17(1):32–40.

Dwinanda, M. W., Satyahadewi, N., and Andani, W. (2023). Classification of Student Graduation Status Using XGBoost ALgorithm. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 17(3):1785–1794.

Fauziningrum, E. and Suryaningsih, E. I. (2021). Evaluasi Dan Prediksi Penguasaan <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/>

- Bahasa Inggris Maritim Menggunakan Metode Decision Tree Dan Confusion Matrix (Studi Kasus Di Universitas Maritim Amni). *Angewandte Chemie International Edition*, 6(11), 951–952., pages 5–24.
- Hann, A., Walter, B. M., and Meining, A. Endoscopic Diagnosis of Celiac Disease.
- Hui, Y., Shuli, L., Rongxiu, L., and Jianyong, Z. (2018). Prediction of component content in rare earth extraction process based on ESNs-Adaboost. *IFAC-PapersOnLine*, 51(21):42–47.
- Irsyad, H. and Taqwiyim, A. (2022). Analisis Ulasan Pengguna Aplikasi Diagnosa Tanaman Di Play Store Menggunakan Naïve Bayes. *Buletin Ilmiah Informatika Teknologi*, 1(2):43–51.
- Jange, B. (2021). Prediksi Harga Saham Bank BCA Menggunakan Prophet. *Journal of Trends Economics and Accounting ...*, 2(1):1–5.
- Jayadi, A. and Meilinda, D. (2023). Klasifikasi Tingkat Kematangan Buah Pepaya Berdasarkan Warna Kulit Menggunakan Sensor Warna TCS3200. *Jurnal ICTEE*, 3(2):1–13.
- Jun-Ho, Tran Mangala, J. J. (2023). Celiac Disease.
- Kara, A., Demirci, E., and Ozmen, S. (2019). Evaluation of psychopathology and quality of life in children with celiac disease and their parents. *Gazi Medical Journal*, 30(1):43–47.
- Katriani, N. and Mailoa, E. (2022). Klasifikasi Bahasa Daerah Menggunakan Decision Tree Dan Gradient Boots. *Jurnal Teknik Informatika dan Sistem Informasi*, 9(2):2407–4322.

- Ketut, I., Jayawardhana, W., Bagus, N., and Kresnapati, A. (2022). Anemia Megaloblastik: Sebuah Tinjauan Pustaka. *BIOCITY Journal of Pharmacy Bioscience and Clinical Community*, 1(1):25–35.
- Kurnia, D., Mazdadi, M. I., Kartini, D., Nugroho, R. A., and Abadi, F. (2023). Seleksi fitur dengan particle swarm optimization pada klasifikasi penyakit parkinson menggunakan xgboost. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 10(5):1083–1094.
- Le, T. T. H., Oktian, Y. E., and Kim, H. (2022). XGBoost for Imbalanced Multiclass Classification-Based Industrial Internet of Things Intrusion Detection Systems. *Sustainability (Switzerland)*, 14(14):1–21.
- Liew, X. Y., Hameed, N., and Clos, J. (2021). An investigation of XGBoost-based algorithm for breast cancer classification. *Machine Learning with Applications*, 6(August):100154.
- Lonang, S. and Normawati, D. (2022). Klasifikasi Status Stunting Pada Balita Menggunakan K-Nearest Neighbor Dengan Feature Selection Backward Elimination. *Jurnal Media Informatika Budidarma*, 6(1):49.
- Luxmi, K. M., Tiwari, N. K., and Ranjan, S. (2023). Estimation and comparison of gabion weir oxygen mass transfer by ensemble learnings of bagging, boosting, and stacking algorithms. *ISH Journal of Hydraulic Engineering*, 0(0):1–16.
- Makharia, G. K., Singh, P., Catassi, C., Sanders, D. S., Leffler, D., Ali, R. A. R., and Bai, J. C. (2022). The global burden of coeliac disease: opportunities and challenges. *Nature Reviews Gastroenterology & Hepatology*, 19:313–327.

- Malik, S., Harode, R., and Singh Kunwar, A. (2020). XGBoost: a deep dive into boosting. *Simon Fraser University*, (February):1–21.
- Melvin, J. and Soraya, A. (2023). Analisis Perbandingan Algoritma XGBoost dan Algoritma Random Forest Ensemble Learning pada Klasifikasi Keputusan Kredit. *Jurnal Riset Rumpun Matematika dan Ilmu Pengetahuan Alam (JURRIMIPA)*, 2(2):87–103.
- Mu'tashim, M. L. and Zaidiah, A. (2023). Klasifikasi Ketepatan Lama Studi Mahasiswa Dengan Algoritma Random Forest Dan Gradient Boosting (Studi Kasus Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta). *Seminar Nasional Mahasiswa Bidang Ilmu Komputer dan Aplikasinya (SENAMIKA)*, 4(1):155–166.
- Nasrullah, A. H. (2021). Implementasi algoritma decision tree untuk klasifikasi produk laris. *Jurnal Ilmiah Ilmu Komputer Fakultas Ilmu Komputer Universitas Al Asyariah Mandar*, 7(2):45–51.
- Nasution, M. K., Widartha, V. P., Telkom, U., and Bayes, N. (2021). Perbandingan Akurasi Algoritma Naive Bayes dan Algoritma XGBoost pada Klasifikasi Penyakit Diabetes. *e-Proceeding of Engineering*, 8(5):9765–9772.
- Niland, B. and Cash, B. D. (2018). Health benefits and adverse effects of a gluten-free diet in non-celiac disease patients. *Gastroenterology and Hepatology*, 14(2):82–91.
- Nuraliza, H., Nurul Pratiwi, O., and Hamami, F. (2022). Analisis Sentimen IMBd Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai Manajemen*, 7(1):1–17.

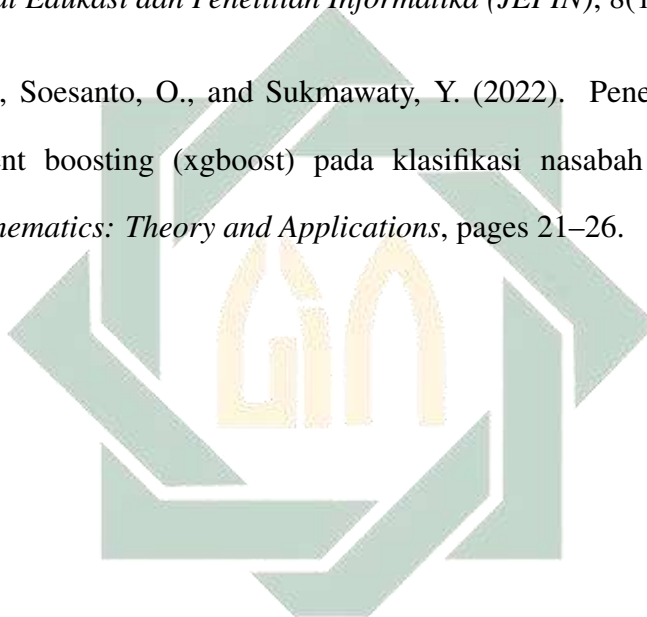
- Oktadiana, H., Abdullah, M., Renaldi, K., and Dyah, N. (2017). Diagnosis dan Tata Laksana Penyakit Celiac. *Jurnal Penyakit Dalam Indonesia*, 4(3):157.
- Orjiekwe, O. A. (2023). Nutritional management of the celiac disease. *Journal of Clinical & Experimental Immunology*, 8(2):561–572.
- Rombe, Y., Thamrin, S. A., and Lawi, A. (2022). Application of Adaptive Synthetic Nominal and Extreme Gradient Boosting Methods in Determining Factors Affecting Obesity: A Case Study of Indonesian Basic Health Research Survey 2013. *Indonesian Journal of Statistics and Its Applications*, 6(2):309–317.
- Sagi, O. and Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information Sciences*, 572:522–542.
- Salsabila, K., Ansori, M., and Paramita, O. (2019). Eksperimen Pembuatan Cupcake Free Gluten Berbahan Dasar Tepung Biji Kluwih dengan Campuran Tepung Beras. *TEKNOBUGA: Jurnal Teknologi Busana dan Boga*, 7(1):31–38.
- Shi, X., Wong, Y. D., Li, M. Z.-F., Palanisamy, C., and Chai, C. (2019). A feature learning approach based on xgboost for driving assessment and risk prediction. *Accident Analysis & Prevention*, 129:170–179.
- Sibindi, R., Mwangi, R. W., and Waititu, A. G. (2023). A boosting ensemble learning based hybrid light gradient boosting machine and extreme gradient boosting model for predicting house prices. *Engineering Reports*, 5(4):1–19.
- Siringoringo, R., Jamaluddin, J., and ... (2022). Text Mining Dan Klasifikasi Multi Label Menggunakan Xgboost. ... : *Jurnal Manajemen ...*, 6(2):234–238.

- Siska, S., Saputra, G. A., Rohmat, C. L., and Sidik, F. (2023). Implementasi Metode Naive Bayes pada Prediksi Penyakit Seliak. *KOPERTIP Jurnal Ilmiah Manajemen Informatika dan Komputer*, 7(1):8–13.
- Stoleru, C. A., Dulf, E. H., and Ciobanu, L. (2022). Automated detection of celiac disease using Machine Learning Algorithms. *Scientific Reports*, 12(1):1–19.
- Sulaiman, T. W., Fitriansyah, R. B., Alaudin, A. R., Ratsanjani, M. H., and Malang, P. N. (2023). Literature Review: Penerapan Big Data dalam Kesehatan Masyarakat. *SATUKATA*, 1(36):129–138.
- Supriadi, M. R. and Andarsyah, R. (2023). *DETEKSI HALAMAN WEBSITE PHISHING MENGGUNAKAN ALGORITMA MACHINE LEARNING GRADIENT BOOSTING CLASSIFIER*. Penerbit Buku Pedia.
- Suyanto (2019). *Data Mining : Untuk Klasifikasi dan Klusterisasi Data*. Penerbit Informatika Bandung, Bandung.
- Syahrizal, H. and Jailani, M. S. (2023). Jenis-Jenis Penelitian Dalam Penelitian Kuantitatif dan Kualitatif. *QOSIM : Jurnal Pendidikan, Sosial & Humaniora*, 1(1):13–23.
- Syihabuddin Azmil Umri, S. (2021). Analisis Dan Komparasi Algoritma Klasifikasi Dalam Indeks Pencemaran Udara Di Dki Jakarta. *JIKO (Jurnal Informatika dan Komputer)*, 4(2):98–104.
- Tim Medis Siloam Hospitals (2023). Penyakit Celiac - Penyebab, Gejala, dan Cara Mengobati.
- Van Der Fels-Klerx, H. J., Smits, N. G., Bremer, M. G., Schultink, J. M., Nijkamp, M. M., Castenmiller, J. J., and De Vries, J. H. (2021). Detection of gluten in <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/> <http://digilib.uinsa.ac.id/>

duplicate portions to determine gluten intake of coeliac disease patients on a gluten-free diet. *British Journal of Nutrition*, 125(9):1051–1057.

Yuhana, U. L. and Purwarianti, A. (2022). Tuning Hyperparameter pada Gradient Boosting. *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, 8(1):134–139.

Yulianti, S. E. H., Soesanto, O., and Sukmawaty, Y. (2022). Penerapan metode extreme gradient boosting (xgboost) pada klasifikasi nasabah kartu kredit. *Journal of Mathematics: Theory and Applications*, pages 21–26.



UIN SUNAN AMPEL
S U R A B A Y A