

**PENERAPAN *ISOLATION FOREST* SEBAGAI DETEKSI *OUTLIER*  
UNTUK KLASIFIKASI PENYAKIT MENGGUNAKAN METODE  
XGBOOST**

**SKRIPSI**



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

Disusun Oleh  
**ARINA NIHAYATA HUSNA**  
**09030222048**

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

**2026**

## PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : ARINA NIHAYATA HUSNA  
NIM : 09030222048  
Program Studi : Matematika  
Angkatan : 2022

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "PENERAPAN *ISOLATION FOREST* SEBAGAI DETEKSI *OUTLIER* UNTUK KLASIFIKASI PENYAKIT MENGGUNAKAN METODE XGBOOST". 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, 29 Desember 2025

Yang menyatakan,



ARINA NIHAYATA HUSNA  
NIM. 09030222048

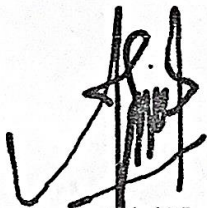
## LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh :

Nama : ARINA NIHAYATA HUSNA  
NIM : 09030222048  
Judul skripsi : PENERAPAN *ISOLATION FOREST* SEBAGAI  
DETEKSI *OUTLIER* UNTUK KLASIFIKASI  
PENYAKIT *AUTOIMUN* MENGGUNAKAN  
METODE XGBOOST

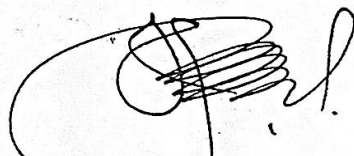
telah diperiksa dan disetujui untuk diujikan.

Pembimbing I



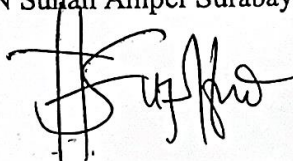
Dr. Dian Candra Rini Novitasari, M. Kom  
NIP. 198511242014032001

Pembimbing II



Dr. Lutfi Hakim, M. Ag.  
NIP. 197312252006041001

Mengetahui,  
Ketua Program Studi Matematika  
UIN Sunan Ampel Surabaya



Dr. Yuniar Farida, M.T  
NIP. 197905272014032002

## PENGESAHAN TIM PENGUJI SKRIPSI

Skripsi oleh:

Nama : ARINA NIHAYATA HUSNA

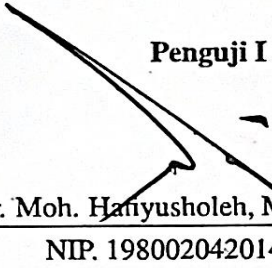
NIM : 09030222048

Judul : PENERAPAN *ISOLATION FOREST* SEBAGAI  
DETEKSI *OUTLIER* UNTUK KLASIFIKASI  
PENYAKIT MENGGUNAKAN METODE XGBOOST

Telah dipertahankan di depan Tim Penguji pada tanggal 5 Januari 2026.

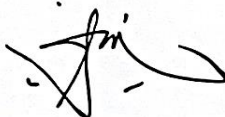
Mengesahkan,  
Tim Penguji

Penguji I

  
Dr. Moh. Haniyusholeh, M.Si., M.PMat.

NIP. 198002042014031001

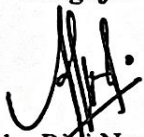
Penguji II



Dian Yuliati, M.Si.

NIP. 198707142020122015

Penguji III



Dr. Dian Candra Rini Novitasari, M. Kom.

NIP. 198511242014032001

Penguji IV



Dr. Lutfi Hakim, M. Ag.

NIP. 197312252006041001

Mengetahui,

Dekan Fakultas Sains dan Teknologi  
UIN Sunan Ampel Surabaya

  
Dr. A. Saepul Hamdani, M.Pd.  
NIP. 196507312000031002

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

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

Nama : Arina Mihayata Husna  
NIM : 09030222048  
Fakultas/Jurusan : Sains dan Teknologi / Matematika  
E-mail address : arinnihayaa00@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 ISOLATION FOREST SEBAGAI DETEKSI  
OUTLIER UNTUK KLASIFIKASI PENYAKIT AUTOIMUN  
MENGUNAKAN METODE XGBOOST

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, 13 Januari 2026

Penulis

  
(Arina Mihayata H.)  
nama terang dan tanda tangan

## ABSTRAK

### **PENERAPAN *ISOLATION FOREST* SEBAGAI DETEKSI *OUTLIER* UNTUK KLASIFIKASI PENYAKIT MENGGUNAKAN METODE XGBOOST**

Penyakit *autoimun* merupakan kondisi ketika sistem kekebalan tubuh menyerang jaringan tubuh sendiri. Keragaman gejala klinis serta tingginya variasi data pasien membuat proses diagnosis menjadi sulit jika hanya mengandalkan metode konvensional. Oleh karena itu, diperlukan pendekatan berbasis *machine learning* yang mampu menangani kompleksitas data dan meningkatkan akurasi klasifikasi. Penelitian ini bertujuan untuk mendeteksi data *outlier* pada data pasien penyakit *autoimun* menggunakan metode *Isolation Forest* serta mengevaluasi kinerja klasifikasi penyakit *autoimun* dengan metode *Extreme Gradient Boosting* (XGBoost) setelah dilakukan penanganan *outlier* dan optimasi parameter. Dataset yang digunakan terdiri dari tujuh kelas penyakit *autoimun* dengan sejumlah variabel klinis sebagai fitur masukan. Proses pengujian dilakukan dengan memvariasikan nilai *contamination percentage* (CP) pada *Isolation Forest*, dilanjutkan dengan *hyperparameter tuning* pada model XGBoost. Evaluasi model dilakukan menggunakan skema *K-Fold Cross Validation* dengan metrik akurasi, sensitivitas, dan spesifisitas. Hasil penelitian menunjukkan bahwa penerapan *Isolation Forest* dengan nilai CP sebesar 10% mampu mengurangi pengaruh data *outlier* dan menghasilkan kinerja klasifikasi yang lebih stabil. Model XGBoost terbaik menghasilkan akurasi sebesar 89.31%, sensitivitas 88.14%, dan spesifisitas 98.18%. Hasil ini menunjukkan keseimbangan yang baik antara kemampuan mendeteksi penyakit dan meminimalkan kesalahan klasifikasi. Dengan demikian, kombinasi *Isolation Forest* dan XGBoost berpotensi dikembangkan sebagai sistem pendukung keputusan untuk membantu tenaga medis dalam diagnosis penyakit *autoimun*.

**Kata kunci:** Penyakit *Autoimun*, klasifikasi, *Isolation Forest*, deteksi *outlier*, XGBoost.

## ABSTRACT

### **Application of Isolation Forest in Outlier Detection for Multiclass Classification of Autoimmune Diseases Using the XGBoost Method**

Autoimmune disease is a condition where the immune system attacks the body's own tissues. The diversity of clinical symptoms and high variation in patient data make the diagnostic process difficult when relying solely on conventional methods. Therefore, a machine learning-based approach is required to handle data complexity and improve classification accuracy. This study aims to detect outliers in autoimmune disease patient data using the Isolation Forest method and evaluate the performance of autoimmune disease classification using the Extreme Gradient Boosting (XGBoost) method after outlier handling and parameter optimization. The dataset consists of seven classes of autoimmune diseases with several clinical variables as input features. The testing process was conducted by varying the contamination percentage (CP) values in Isolation Forest, followed by hyperparameter tuning on the XGBoost model. Model evaluation was performed using a K-Fold Cross Validation scheme with accuracy, sensitivity, and specificity as metrics. The results showed that the application of Isolation Forest with a CP value of 10% effectively reduced the influence of outliers and produced more stable classification performance. The best XGBoost model achieved an accuracy of 89.31%, sensitivity of 88.14%, and specificity of 98.18%. These results demonstrate a good balance between the ability to detect the disease and minimizing misclassification. Thus, the combination of Isolation Forest and XGBoost has the potential to be developed as a decision support system to assist medical professionals in diagnosing autoimmune diseases.

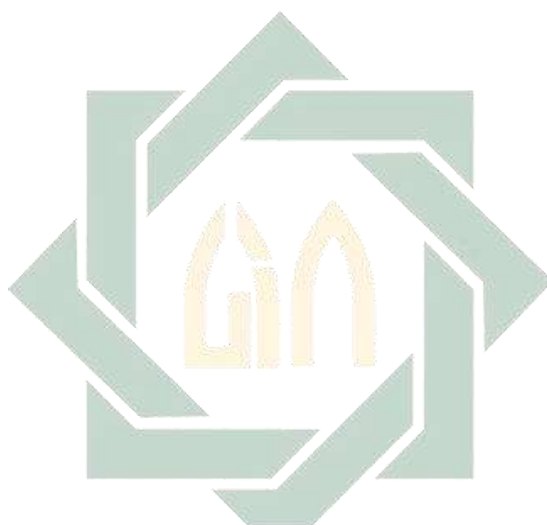
**Keywords:** Autoimmune Disease, Classification, *Isolation Forest*, Outlier Detection, *XGBoost*.

# DAFTAR ISI

|                                                            |          |
|------------------------------------------------------------|----------|
| HALAMAN JUDUL                                              | i        |
| PENGESAHAN TIM PENGUJI SKRIPSI                             | ii       |
| HALAMAN PERNYATAAN KEASLIAN                                | iii      |
| MOTTO                                                      | iv       |
| PERSEMBAHAN                                                | v        |
| KATA PENGANTAR                                             | vi       |
| DAFTAR ISI                                                 | viii     |
| DAFTAR TABEL                                               | xi       |
| DAFTAR GAMBAR                                              | xiii     |
| ABSTRAK                                                    | xvi      |
| ABSTRACT                                                   | xvii     |
| <b>I PENDAHULUAN</b>                                       | <b>1</b> |
| 1.1. Latar Belakang Masalah . . . . .                      | 1        |
| 1.2. Rumusan Masalah . . . . .                             | 6        |
| 1.3. Tujuan Penelitian . . . . .                           | 6        |
| 1.4. Manfaat Penelitian . . . . .                          | 6        |
| 1.5. Batasan Masalah . . . . .                             | 7        |
| 1.6. Sistematika Penulisan . . . . .                       | 7        |
| <b>II TINJAUAN PUSTAKA</b>                                 | <b>9</b> |
| 2.1. Penyakit <i>Autoimun</i> . . . . .                    | 9        |
| 2.1.1. <i>Rheumatoid Arthritis (RA)</i> . . . . .          | 10       |
| 2.1.2. <i>Reactive Arthritis (ReA)</i> . . . . .           | 11       |
| 2.1.3. <i>Ankylosing Spondylitis (AS)</i> . . . . .        | 12       |
| 2.1.4. <i>Sjögren's Syndrome (SS)</i> . . . . .            | 13       |
| 2.1.5. <i>Systemic Lupus Erythematosus (SLE)</i> . . . . . | 14       |
| 2.1.6. <i>Psoriatic Arthritis (PA)</i> . . . . .           | 14       |
| 2.2. <i>Isolation Forest (IF)</i> . . . . .                | 15       |

|                                                                                                                                     |            |
|-------------------------------------------------------------------------------------------------------------------------------------|------------|
| 2.3. <i>Adaptive Synthetic Sampling</i> (ADASYN) . . . . .                                                                          | 19         |
| 2.4. <i>K-Fold Cross Validation</i> . . . . .                                                                                       | 22         |
| 2.5. <i>Extreme Gradient Boosting</i> (XGBoost) . . . . .                                                                           | 24         |
| 2.6. <i>Confusion Matrix</i> . . . . .                                                                                              | 29         |
| 2.7. Integrasi Keislaman . . . . .                                                                                                  | 32         |
| <b>III METODE PENELITIAN</b>                                                                                                        | <b>38</b>  |
| 3.1. Jenis Penelitian . . . . .                                                                                                     | 38         |
| 3.2. Sumber Data . . . . .                                                                                                          | 38         |
| 3.3. Variabel Penelitian . . . . .                                                                                                  | 38         |
| 3.3.1. Variabel fitur . . . . .                                                                                                     | 39         |
| 3.3.2. Variabel kelas . . . . .                                                                                                     | 40         |
| 3.4. Metode Analisis Data . . . . .                                                                                                 | 41         |
| 3.4.1. Input Dataset . . . . .                                                                                                      | 41         |
| 3.4.2. Penanganan Data <i>Imbalance</i> . . . . .                                                                                   | 41         |
| 3.4.3. Deteksi data <i>outlier</i> . . . . .                                                                                        | 42         |
| 3.4.4. Klasifikasi . . . . .                                                                                                        | 43         |
| 3.4.5. Evaluasi Hasil . . . . .                                                                                                     | 46         |
| 3.5. Skema Uji Coba <i>Tuning Parameter</i> . . . . .                                                                               | 46         |
| <b>IV HASIL DAN PEMBAHASAN</b>                                                                                                      | <b>47</b>  |
| 4.1. <i>Pre-processing</i> . . . . .                                                                                                | 47         |
| 4.1.1. Penanganan <i>Outlier</i> menggunakan <i>Isolation Forest</i> . . . . .                                                      | 48         |
| 4.1.2. Penanganan <i>Data Imbalance</i> menggunakan ADASYN . . . . .                                                                | 73         |
| 4.2. Klasifikasi . . . . .                                                                                                          | 78         |
| 4.2.1. Proses <i>Training</i> . . . . .                                                                                             | 78         |
| 4.2.2. Pohon Pertama . . . . .                                                                                                      | 80         |
| 4.2.3. Pohon Kedua . . . . .                                                                                                        | 90         |
| 4.2.4. Proses <i>Testing</i> . . . . .                                                                                              | 113        |
| 4.3. Pengujian Model dan Evaluasi Hasil . . . . .                                                                                   | 118        |
| 4.4. Analisis Pengaruh Nilai <i>Contamination Percentage</i> (CP)<br>Isolation Forest terhadap Kinerja Model pada $K = 7$ . . . . . | 132        |
| 4.5. Aplikasi Deteksi Jenis Penyakit <i>Autoimun</i> . . . . .                                                                      | 139        |
| 4.6. Integrasi Keislaman . . . . .                                                                                                  | 140        |
| <b>V PENUTUP</b>                                                                                                                    | <b>143</b> |
| 5.1. Kesimpulan . . . . .                                                                                                           | 143        |

|                       |            |
|-----------------------|------------|
| 5.2. Saran . . . . .  | 144        |
| <b>DAFTAR PUSTAKA</b> | <b>161</b> |



UIN SUNAN AMPEL  
S U R A B A Y A

# DAFTAR TABEL

|      |                                                                                       |     |
|------|---------------------------------------------------------------------------------------|-----|
| 3.1  | Sampel Data Penyakit <i>Autoimun</i> . . . . .                                        | 39  |
| 3.2  | Parameter dan Nilai yang Digunakan dalam Penelitian . . . .                           | 46  |
| 4.1  | Sampel <i>Dataset</i> . . . . .                                                       | 47  |
| 4.2  | Hasil Kedalaman <i>Isolation Forest</i> . . . . .                                     | 68  |
| 4.3  | Rata-Rata <i>Path Length</i> Pada Masing-Masing Data . . . . .                        | 69  |
| 4.4  | Hasil Skor <i>Outlier</i> . . . . .                                                   | 71  |
| 4.5  | Pelabelan Skor <i>Outlier</i> Pada Data . . . . .                                     | 72  |
| 4.6  | Jumlah Data Training dan Data Testing . . . . .                                       | 78  |
| 4.7  | Data Pada Proses <i>Training</i> . . . . .                                            | 79  |
| 4.8  | Contoh Data Training pada Pohon Pertama . . . . .                                     | 81  |
| 4.9  | Nilai <i>Gain Node</i> ke-1 Pohon Pertama Kelas <i>Rheumatoid Arthritis</i> . . . . . | 83  |
| 4.10 | Nilai <i>Gain Node</i> ke-2 Pohon Pertama Kelas <i>Rheumatoid Arthritis</i> . . . . . | 84  |
| 4.11 | Nilai <i>Gain Node</i> ke-3 Pohon Pertama Kelas <i>Rheumatoid Arthritis</i> . . . . . | 85  |
| 4.12 | Nilai <i>Gain Node</i> ke-4 Pohon Pertama Kelas <i>Rheumatoid Arthritis</i> . . . . . | 86  |
| 4.13 | Perhitungan Residual Pada Pohon Pertama Kelas <i>Psoriatic Arthritis</i> . . . . .    | 87  |
| 4.14 | Nilai <i>Leaf</i> Pada Pohon Pertama untuk Masing-Masing Kelas                        | 88  |
| 4.15 | Data Pada Proses <i>Testing</i> . . . . .                                             | 113 |
| 4.16 | Nilai <i>Leaf</i> Data <i>Testing</i> . . . . .                                       | 114 |
| 4.17 | Nilai Probabilitas Masing-Masing Kelas Sampel Data <i>Testing</i>                     | 117 |
| 4.18 | Hasil Evaluasi Model pada Fold 5 - 10 . . . . .                                       | 119 |
| 4.19 | Hasil Uji Coba CP=5% . . . . .                                                        | 120 |

|                                                                                               |     |
|-----------------------------------------------------------------------------------------------|-----|
| 4.20 Hasil Uji Coba CP=6% . . . . .                                                           | 121 |
| 4.21 Hasil Uji Coba CP=7% . . . . .                                                           | 122 |
| 4.22 Hasil Uji Coba CP=8% . . . . .                                                           | 124 |
| 4.23 Hasil Uji Coba CP = 9% . . . . .                                                         | 125 |
| 4.24 Hasil Uji Coba CP = 10% . . . . .                                                        | 126 |
| 4.25 Hasil Uji Coba Setiap <i>Fold</i> Pada CP=10% . . . . .                                  | 128 |
| 4.26 Confusion Matrix pada Setiap Fold . . . . .                                              | 129 |
| 4.27 Perbandingan Statistik Seluruh Fitur Sebelum dan Sesudah<br>Penanganan Outlier . . . . . | 133 |
| 4.28 Perbandingan Hasil Kinerja Dengan Penelitian Lainnya . . .                               | 138 |



UIN SUNAN AMPEL  
S U R A B A Y A

# DAFTAR GAMBAR

|      |                                                                                                                        |    |
|------|------------------------------------------------------------------------------------------------------------------------|----|
| 2.1  | Perbandingan Sendi Normal dan Sendi Rheumatoid Arthritis                                                               | 11 |
| 2.2  | Ilustrasi penderita Reactive Arthritis . . . . .                                                                       | 11 |
| 2.3  | Perbandingan tulang belakang penderita <i>Ankylosing Spondylitis</i> dan Normal . . . . .                              | 12 |
| 2.4  | Gejala penderita <i>Sjögren's Syndrome</i> . . . . .                                                                   | 13 |
| 2.5  | Kulit penderita <i>Systemic Lupus Erythematosus (SLE)</i> . . . . .                                                    | 14 |
| 2.6  | Ilustrasi penderita Psoriatic Arthritis . . . . .                                                                      | 15 |
| 2.7  | Arsitektur <i>Isolation Forest</i> . . . . .                                                                           | 16 |
| 2.8  | Cara Kerja <i>Isolation Forest</i> . . . . .                                                                           | 19 |
| 2.9  | Ilustrasi cara kerja ADASYN: (a) pembangkitan data sintetis, (b) distribusi data setelah <i>oversampling</i> . . . . . | 20 |
| 2.10 | Skema Proses <i>K-Fold Cross Validation</i> . . . . .                                                                  | 23 |
| 2.11 | Arsitektur Pada XGBoost . . . . .                                                                                      | 25 |
| 2.12 | Ilustrasi Umum <i>Confusion Matrix</i> . . . . .                                                                       | 29 |
| 2.13 | Hasil CM Tiap Kelas pada Klasifikasi Penyakit <i>Autoimun</i> . . . . .                                                | 31 |
| 3.1  | Diagram Alir Penelitian . . . . .                                                                                      | 41 |
| 3.2  | Diagram Alir <i>Isolation Forest</i> (IF) . . . . .                                                                    | 43 |
| 3.3  | Diagram Alir <i>Training</i> XGBoost . . . . .                                                                         | 44 |
| 3.4  | Diagram Alir <i>Testing</i> XGBoost . . . . .                                                                          | 45 |
| 4.1  | Visualisasi Data <i>Outlier</i> Sebelum Penanganan . . . . .                                                           | 49 |
| 4.2  | Pohon ke-1 <i>Isolation Forest</i> . . . . .                                                                           | 52 |
| 4.3  | <i>Path</i> ke-1 Pohon 1 <i>Isolation Forest</i> . . . . .                                                             | 53 |
| 4.4  | <i>Path</i> ke-2 Pohon 1 <i>Isolation Forest</i> . . . . .                                                             | 54 |
| 4.5  | <i>Path</i> ke-3 Pohon 1 <i>Isolation Forest</i> . . . . .                                                             | 55 |
| 4.6  | <i>Path</i> ke-4 Pohon 1 <i>Isolation Forest</i> . . . . .                                                             | 56 |
| 4.7  | <i>Path</i> ke-5 Pohon 1 <i>Isolation Forest</i> . . . . .                                                             | 57 |

|      |                                                                     |     |
|------|---------------------------------------------------------------------|-----|
| 4.8  | <i>Path ke-6 Pohon 1 Isolation Forest</i>                           | 58  |
| 4.9  | <i>Path ke-7 Pohon 1 Isolation Forest</i>                           | 59  |
| 4.10 | <i>Path ke-8 Pohon 1 Isolation Forest</i>                           | 60  |
| 4.11 | <i>Path ke-9 Pohon 1 Isolation Forest</i>                           | 61  |
| 4.12 | <i>Path ke-10 Pohon 1 Isolation Forest</i>                          | 62  |
| 4.13 | <i>Path ke-11 Pohon 1 Isolation Forest</i>                          | 63  |
| 4.14 | <i>Path ke-12 Pohon 1 Isolation Forest</i>                          | 64  |
| 4.15 | <i>Pohon ke-2 Isolation Forest dengan kedalaman 5</i>               | 65  |
| 4.16 | <i>Pohon ke-3 Isolation Forest dengan kedalaman 5</i>               | 66  |
| 4.17 | <i>Pohon ke-100 Isolation Forest dengan kedalaman 5</i>             | 67  |
| 4.18 | <i>Pohon Pada Node ke-1 Split</i>                                   | 81  |
| 4.19 | <i>Pohon Pada Node ke-1 Split</i>                                   | 82  |
| 4.20 | <i>Pohon Pada Node ke-1 Split</i>                                   | 83  |
| 4.21 | <i>Pohon pada Node ke-2</i>                                         | 84  |
| 4.22 | <i>Pohon pada Node ke-3 dengan Split</i>                            | 85  |
| 4.23 | <i>Pohon pada Node ke-4</i>                                         | 87  |
| 4.24 | <i>Pohon Keputusan Ke-1 Pada Kelas Ankylosing Spondylitis</i>       | 93  |
| 4.25 | <i>Pohon Keputusan Ke-1 Pada Kelas Normal</i>                       | 94  |
| 4.26 | <i>Pohon Keputusan Ke-1 Pada Kelas Psoriatic Arthritis</i>          | 95  |
| 4.27 | <i>Pohon Keputusan Ke-1 Pada Kelas Reactive Arthritis</i>           | 96  |
| 4.28 | <i>Pohon Keputusan Ke-1 Pada Kelas Rheumatoid Arthritis</i>         | 97  |
| 4.29 | <i>Pohon Keputusan Ke-1 Pada Kelas Sjögren's Syndrome</i>           | 98  |
| 4.30 | <i>Pohon Keputusan Ke-1 Pada Kelas Systemic Lupus Erythematosus</i> | 99  |
| 4.31 | <i>Pohon Keputusan Ke-2 Pada Kelas Ankylosing Spondylitis</i>       | 100 |
| 4.32 | <i>Pohon Keputusan Ke-2 Pada Kelas Normal</i>                       | 101 |
| 4.33 | <i>Pohon Keputusan Ke-2 Pada Kelas Psoriatic Arthritis</i>          | 102 |
| 4.34 | <i>Pohon Keputusan Ke-2 Pada Kelas Reactive Arthritis</i>           | 103 |
| 4.35 | <i>Pohon Keputusan Ke-2 Pada Kelas Rheumatoid Arthritis</i>         | 104 |
| 4.36 | <i>Pohon Keputusan Ke-2 Pada Kelas Sjögren's Syndrome</i>           | 105 |
| 4.37 | <i>Pohon Keputusan Ke-2 Pada Kelas Systemic Lupus Erythematosus</i> | 106 |
| 4.38 | <i>Pohon Keputusan Ke-700 Pada Kelas Ankylosing Spondylitis</i>     | 107 |
| 4.39 | <i>Pohon Keputusan Ke-700 Pada Kelas Normal</i>                     | 108 |

|      |                                                                                 |     |
|------|---------------------------------------------------------------------------------|-----|
| 4.40 | Pohon Keputusan Ke-700 Pada Kelas <i>Psoriatic Arthritis</i> . . .              | 109 |
| 4.41 | Pohon Keputusan Ke-700 Pada Kelas <i>Reactive Arthritis</i> . . .               | 110 |
| 4.42 | Pohon Keputusan Ke-700 Pada Kelas <i>Rheumatoid Arthritis</i> . . .             | 111 |
| 4.43 | Pohon Keputusan Ke-700 Pada Kelas <i>Sjögren's Syndrome</i> . . .               | 112 |
| 4.44 | Pohon Keputusan Ke-700 Pada Kelas <i>Systemic Lupus Erythematosus</i> . . . . . | 113 |
| 4.45 | Visualisasi Hasil Uji Coba Pada Uji Coba Terbaik Setiap CP                      | 127 |
| 4.46 | Visualisasi Hasil Uji Coba CP Dengan Data <i>Inlier</i> . . . . .               | 134 |
| 4.47 | Visualisasi Hasil Uji Coba CP Dengan Imputasi <i>Outlier</i> . . .              | 135 |
| 4.48 | Visualisasi Hasil Penanganan Data <i>Outlier</i> . . . . .                      | 137 |
| 4.49 | Tampilan Utama Aplikasi Deteksi Jenis Penyakit <i>Autoimun</i> .                | 139 |



UIN SUNAN AMPEL  
S U R A B A Y A

## DAFTAR PUSTAKA

- (2024). Mekanisme adaptasi sel tubuh terhadap serangan penyakit autoimun: Lupus. *Jurnal Penelitian Perawat Profesional*, 6(4):1687–1694.
- Abadi, S. N. and colleagues (2025). Comparative analysis of xgboost algorithm and linear regression in predicting the trend of investor overreaction. *Journal of Business, Economics and Management*. The study compared multiple n\_estimators (100–600) and found the optimal configuration at n\_estimators = 400 for best performance.
- Abdurrahman, G., Oktavianto, H., and Sintawati, M. (2022). Optimasi algoritma xgboost classifier menggunakan hyperparameter gridsearch dan random search pada klasifikasi penyakit diabetes. *Informal Informatics J*, 7(3):193.
- Adiputra, B. (2023). Dinamika resiliensi pada orang dengan lupus (odapus). *Jurnal Empati*.
- Aji, B. A. S., Setiawan, Y., and Anggraini, S. D. (2025). Analisis perbandingan algoritma decision tree, random forest, dan xgboost untuk klasifikasi penyakit infeksi gigi dan mulut. *INTEGER: Journal of Information Technology*, 10(1).
- Almuzadid, G. and Subhiyakto, E. R. (2025). Stroke risk classification using the ensemble learning method of xgboost and random forest. *Journal of Applied Informatics and Computing*, 9(3):828–837.
- Alodokter (2022). Rheumatoid arthritis – gejala, penyebab dan mengobati. Diakses tanggal 25 September 2025.
- Amanda, N. D., Nurhidayah, T. M., Ramadhani, T. Y., et al. (2023). Menjaga kesehatan tubuh dalam perspektif islam. *Religion: Jurnal Agama, Sosial, Dan Budaya*, 2(5):373–380.
- Andryan, M. R., Fajri, M., and Sulistyowati, N. (2022). Komparasi kinerja algoritma xgboost dan algoritma support vector machine (svm)

- untuk diagnosa penyakit kanker payudara. *JIKO (Jurnal Informatika dan Komputer)*, 6(1):1–5. Akurasi XGBoost 95.12%, AUC 0.99; SVM 90.24%, AUC 0.98.
- Angkat, C. I., Sianturi, I. M., Sinambela, S. H., Iqbal, M., and Wijaya, R. F. (2024). Sistem pakar diagnosa rheumatoid arthritis dengan menerapkan algoritma teorema bayes. *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, 7(1):08–15.
- Anugrah, M. I., Zeniarja, J., and Setiawan, D. S. (2024). Peningkatan performa model hard voting classifier dengan teknik oversampling adasyn pada dataset diabetes. *Edumatic: Jurnal Pendidikan Informatika*, 8(1):290–299.
- Arfianda, A., Tharida, M., and Masthura, S. (2022). Faktor-faktor yang mempengaruhi terjadinya penyakit rheumatoid arthritis pada lansia di gampong piyeung manee kecamatan montasik kabupaten aceh besar. *JOURNAL OF HEALTHCARE TECHNOLOGY AND MEDICINE*, 8(2):992–1002.
- Arum, D. M. P., Santoso, K. I., Triyono, A., Supriyadi, E., Nugroho, A. S., and Widodo, E. (2025). Algoritma random forest, decision tree, dan xgboost untuk klasifikasi stunting pada balita. *Jurnal Transformatika*, 23(1):67–76.
- Australia, H. (2025). Ankylosing spondylitis – symptoms, diagnosis and treatment. Diakses 28 Oktober 2025.
- Cahyadi, E. (2023). Penyuluhan pengelolaan penyakit autoimun di masyarakat lam bunot, kecamatan simpang tiga. *Jurnal Pengabdian Kepada Masyarakat*, pages 196–201. Diterima 25 Agustus 2023; Disetujui 30 Agustus 2023; Dipublikasi 15 September 2023.
- Centre, A. A. . R. (2025). Psoriatic arthritis – signs/symptoms, diagnosis, treatment. Diakses 28 Oktober 2025.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. pages 785–794.

- Dachi, J. M. A. S. and Sitompul, P. (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.
- Desiani, A., Nurhaliza, S., Putri, T. F., and Suprihatin, B. (2025). Algoritma extreme gradient boosting (xgboost) dan adaptive boosting (adaboost) untuk klasifikasi penyakit tiroid. *Jurnal Rekayasa Elektro Sriwijaya*, 6(2):66–75.
- Diantini, D. M. A., Ulandari, N. L., Wirandani, N. K. N. S., Niruri, R., and Kumara, K. D. (2024). Angka kejadian penyakit autoimun pada pasien anak di rsup sanglah denpasar. *Jurnal Farmasi Udayana*, 5(2):30–34.
- Dwiariputra, Y. H. (2022). *Mendeteksi Anomali Pada Data Transaksi Jual Beli Akta Tanah Dengan Menggunakan Algoritma Decision Tree, Isolation Forest, Dan Random Forest*. PhD thesis, Universitas Mercu Buana Jakarta.
- Firdaus, R. (2025). Classification of autoimmune diseases using the k-nearest neighbor algorithm. *Formosa Journal of Science and Technology*, 5(1):45–54.
- Hakim, A., Sholihah, F. M., and Anifa, N. A. (2023). Konsep ikhtiar dalam berobat sesuai ajaran islam. *Religion: Jurnal Agama, Sosial, Dan Budaya*, 2(4):914–924.
- Hanifah, D. (2025). *Faktor Penyebab Kekambuhan Pada Pasien Rheumatoid Arthritis*. PhD thesis, Universitas Muhammadiyah Malang.
- Harahap, R., Irpan, M., Dinata, M. A., Efrizoni, L., et al. (2024). Perbandingan algoritma random forest dan xgboost untuk klasifikasi penyakit paru-paru berdasarkan data demografi pasien. *BETRIK*, 15(02):130–141.
- Heldayani, I. (2024). Penderita systemic lupus erythematosus (sle). *Jurnal Kesehatan Unpad*, 12(2).

- Hidayanti, D., Wilyo, F. A., Saragih, M., Setyaningrum, O., and Rinaldo, A. (2023). Patofisiologi penyakit meniere pada pasien autoimun: Tinjauan literatur. *Health Information: Jurnal Penelitian*, pages e1061–e1061.
- Ismanda, R. S., Silitonga, M. T. A., and Nur hasanah, S. (2025). Deteksi hybrid anomali transaksi digital dengan optimasi isolation forest-k-means untuk peningkatan keamanan finansial. *Innovative: Journal Of Social Science Research*, 5(3):5749–5765.
- Karim, A., Esabella, S., Kusmanto, K., Suryadi, S., and Purba, E. (2023). Penerapan metode teorema bayes dalam mendiagnosa penyakit autoimun. *Build. Informatics, Technol. Sci*, 5(1):254–263.
- Kemenkes (2022). *Profil Kesehatan Indonesia 2022*. Kementerian Kesehatan RI, Jakarta. Berisi data-kesehatan nasional Indonesia termasuk penyakit autoimun.
- Khine, M. S., Liu, Y., Pallipuram, V. K., and Afari, E. (2024). A machine-learning approach to predicting the achievement of australian students using school climate, learner characteristics, and economic, social, and cultural status. *Education Sciences*, 14(12):1350.
- Kurniawan, A. (2024). Deep learning for autoimmune disease diagnosis and classification. *International Journal of Medical Informatics*.
- Kusuma, B. (2024). Perbandingan metode ensemble machine learning untuk klasifikasi tenaga kerja di indonesia dengan random forest, xgboost, dan catboost. *Jurnal Teknik Sistem dan Komputer*.
- Li, J. et al. (2019). Xgboost for classification and regression using genetic algorithm for hyperparameter optimization. *Neurocomputing*, 367:67–77.
- Mahasiswa, J. (2025). Klasifikasi penyakit autoimun dengan algoritma k-nearest neighbors. *Jurnal Biologi dan Kesehatan*.
- Mahayasa, I. W. et al. (2024). Diagnostic problem in systemic lupus erythematosus: Ana and classification criteria. *Jurnal Biologi Tropis*, 24(1b):296–302.

- Mahdi, M. F., Jahani, A., and Abd, D. H. (2025). Fuzzy evaluation and explainable machine learning for diagnosis of rheumatic and autoimmune diseases. *PeerJ Computer Science*, 11:e3096.
- Markowiak, S. (2024). Ankylosing spondylitis: Review of clinical features, diagnosis, and treatment. *Journal of Environmental Health Science*, 1:51–60.
- Napitupulu, M. (2024). Analisis penyakit dan gejala autoimun menggunakan pola klasifikasi. *Jurnal SNISTIK*.
- Opitasari, O., Natsir, F., and Marsiani, E. S. (2024). Klasifikasi diagnosis untuk penyakit kanker serviks menggunakan algoritma extreme gradient boosting (xgboost). *Jurnal Ilmiah FIFO*, 16(1):55.
- OraCoat (2025). Sjogrens's syndrome: Dry mouth symptoms and treatment. Diakses tanggal 28 Oktober 2025.
- Pamuji, F. (2023). Komparasi metode smote dan adasyn untuk penanganan data tidak seimbang pada klasifikasi multikelas. *Jurnal Informatika Polinema*, 11(3):331–338.
- Pramudya, M. R. A., Arafat, C., Ramadhan, M. C., Jafar, F. A., and Ismanto, E. (2025). A perbandingan algoritma random forest dan xgboost untuk klasifikasi penyakit jantung berdasarkan data medis. *JURNAL FASILKOM*, 15(2):423–428.
- Prasetya, M. G. (2024). *GAMBARAN KADAR C-REACTIVE PROTEIN (CRP) DAN LAJU ENDAP DARAH (LED) PADA PASIEN RHEUMATOID ARTHRITIS DI RSUD DR. H. ABDUL MOELOEK TAHUN 2022-2023*. PhD thesis, POLTEKKES KEMENKES TANJUNGPONOROK.
- Rahmat, F. (2024). Klasifikasi diagnosis untuk penyakit kanker serviks menggunakan algoritma extreme gradient boosting (xgboost). *Jurnal Sistem Informasi FIFO*, 17(1):25–35.
- Rahmawati, A. et al. (2025). A comparative performance of smote, adasyn and random oversampling in machine learning models on prostate cancer dataset. *Jurnal AI Community*, 9(3).

- Rahmawati, F. (2022). Sjogren's syndrome. *Jurnal Ilmiah Kedokteran Wijaya Kusuma*, 1(1):46–57.
- Ramadhan, N. (2021). Sistem pakar diagnosa penyakit autoimun menggunakan metode adasyn dan random forest. *Jurnal Klik*.
- Ramadhanti, D. V., Santoso, R., and Widiharhi, T. (2022). Perbandingan smote dan adasyn pada data imbalance untuk klasifikasi rumah tangga miskin kabupaten temanggung dengan algoritma k-nearest neighbor. *Jurnal Gaussian*, 11(4):499–505.
- Rosyid, S. (2025). *Graph Auto Encoder Untuk Deteksi Outlier Pada Model XGBoost Regression*. PhD thesis, Universitas Gadjah Mada.
- Rosyidah, W. C. (2022). *KEEFEKTIFAN COGNITIVE BEHAVIORAL THERAPY UNTUK MENURUNKAN GEJALA ANSIETAS DAN DEPRESI PASIEN SJOGREN'S SYNDROME*. PhD thesis, UNS (Sebelas Maret University).
- Sadikin, F. (2025). Identifikasi variasi gen yang bersifat missense/nonsense pada penyakit autoimun dermatomiositis. *Jurnal Medika Farmasi*, 12(3):255–265.
- Sajjad, F. S., Handayani, W., Mariam, L., and Mulianingsih, W. (2024). Relationship between stress level, body mass index (bmi), and smoking behavior with severity of psoriasis at hospital x in central lombok region. *Jurnal Biologi Tropis*, 24(1b):631–638.
- Santoso, D. B. and Wahyuni, Y. (2024). Sistem log web server sebagai pendeteksi anomali menggunakan isolation forest. *JUBIKOM: Jurnal Aplikasi Bisnis dan Komputer*, 4(3):90–96.
- Sapari, A. M. (2023). Air quality classification using extreme gradient boosting (xgboost). *Innovation in Research of Informatics*, 5(2):44–51.
- Sarianti, D. and Rini, Y. N. (2023). Penyembuhan berbagai penyakit menurut perspektif islam. *Journal Islamic Education*, 1(3):569–578.
- Setiawan, D., Muhammad, A., and Dewi, S. H. F. (2025). Penerapan algoritma klasifikasi untuk deteksi dini penyakit jantung koroner

- berdasarkan gejala klinis. *Teknik: Jurnal Ilmu Teknik dan Informatika*, 5(1):18–26.
- Sinaga, H. (2022). Perbandingan metode decision tree dan xgboost untuk klasifikasi tingkat penyakit hepatitis c. *Jurnal Teknologi Informasi dan Ilmu Komputer*.
- Soelistijadi, R., Wismarini, T. D., Eniyati, S., and Sunardi, S. (2024). Pemodelan prediktif menggunakan metode ensemble learning xgboost dalam peningkatan akurasi klasifikasi penyakit ginjal. *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)*, 5(4):1866–1875.
- Staff, M. C. (2022). Reactive arthritis — symptoms. Diakses 28 Oktober 2025.
- Supriatna, E. (2021). Wabah corona virus disease covid 19 dalam pandangan islam. *SALAM: Jurnal Sosial Dan Budaya Syar-I*, 7(6):555–564.
- Syafiyati, W. (2025). Mengeksplorasi hubungan antara nutrisi dan hasil penyakit autoimun: Tinjauan sistematis terhadap bukti terkini. *Jurnal Kesehatan Sainika Meditory*, 8(1):259–266.
- Syahputra, M., Defit, S., and Sumijan, S. (2021). Sistem pakar metode case based reasoning untuk mengidentifikasi penyakit psoriasis. *Jurnal Sistim Informasi Dan Teknologi*, pages 20–27.
- Syamsu, R. F., Nawawi, N. M., and Hazminafni, N. N. K. (2025). Studi literature: Potensi resveratrol sebagai efek immunodulator dalam pencegahan penyakit autoimun. *Malahayati Nursing Journal*, 7(8).
- Syukron, M., Santoso, R., and Widiharih, T. (2021). Perbandingan metode smote random forest dan smote xgboost untuk klasifikasi tingkat penyakit hepatitis c pada imbalance class data. *Jurnal Gaussian*, 9(3):227–236.
- Tjan, B. (2022). Gambaran profil systemic lupus erythematosus (sle) dan karakteristik pasien. *Jurnal Penelitian dan Pengembangan*.

- Triana, O. (2024). Deteksi anomali jaringan menggunakan algoritma isolation forest. *Jurnal Dunia Data*, 1(5).
- Wibawa, I. M. S. T. and Karyawati, A. A. I. N. E. (2023). Isolation forest dengan exploratory data analysis pada anomaly detection untuk data transaksi. *JNATIA: Jurnal Nasional Teknologi Informasi dan Aplikasinya*, 1(3):803–809.
- Wikipedia (2025). Lupus. Diakses 28 Oktober 2025.
- Yang, D., Peng, X., Zheng, S., and Peng, S. (2025). Deep learning-based prediction of autoimmune diseases. *Scientific Reports*, 15(1).
- Yifeng, D. (2024). Study on breast cancer classification prediction based on xgboost. *IEEE Access*.
- Zhu, W., Shen, S., and Zhang, Z. (2022). Improved multiclassification of schizophrenia based on xgboost and information fusion for small datasets. *Computational and Mathematical Methods in Medicine*, 2022:1–11.
- Zizilia, R., Chrisnanto, Y. H., and Abdillah, G. (2025). Lung cancer classification using the extreme gradient boosting (xgboost) algorithm and mutual information for feature selection. *SISTEMASI*, 14(5):2198–2214.

UIN SUNAN AMPEL  
S U R A B A Y A