

**IMPLEMENTASI ARSITEKTUR BACKEND FASTAPI
DAN API GATEWAY UNTUK INTEGRASI MULTIPLE
MODEL MACHINE LEARNING**

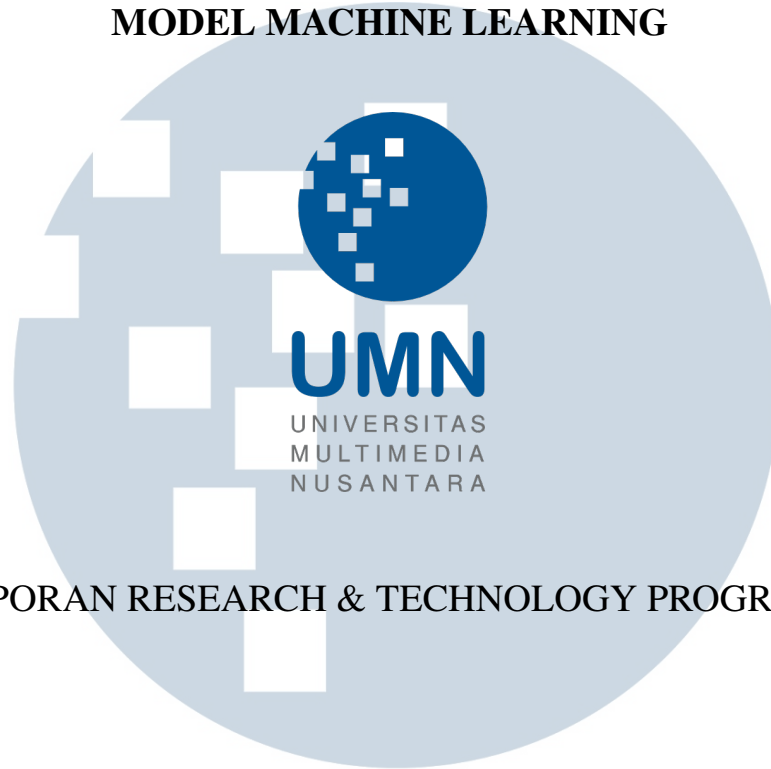


LAPORAN RESEARCH & TECHNOLOGY PROGRAM

**JACQUES FARRELL DHARMA
00000069027**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

**IMPLEMENTASI ARSITEKTUR BACKEND FASTAPI
DAN API GATEWAY UNTUK INTEGRASI MULTIPLE
MODEL MACHINE LEARNING**



LAPORAN RESEARCH & TECHNOLOGY PROGRAM

**JACQUES FARRELL DHARMA
00000069027**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2026**

HALAMAN PERNYATAAN ORISINALITAS TIDAK PLAGIAT

Dengan ini saya,

Nama : Jacques Farrell Dharma

NIM : 00000069027

Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Laporan Research & Technology Program saya yang berjudul:

Implementasi Arsitektur Backend FastAPI dan API Gateway untuk Integrasi Multiple Model Machine Learning

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 20 Desember 2025



(Jacques Farrell Dharma)

HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Jacques Farrell Dharma
NIM : 00000069027
Program Studi : Informatika
Judul Laporan : Implementasi Arsitektur Backend FastAPI dan API Gateway
untuk Integrasi Multiple Model Machine Learning

Dengan ini saya menyatakan secara jujur telah menggunakan bantuan Kecerdasan Artifisial (AI) dalam pengerjaan Laporan dengan rincian sebagai berikut:

- ✓ Menggunakan AI sebagaimana diizinkan untuk membantu dalam menghasilkan ide-ide utama saja.
- ☐ Menggunakan AI sebagaimana diizinkan untuk membantu menghasilkan teks pertama saja.
- ✓ Menggunakan AI untuk menyempurnakan sintaksis dan tata bahasa untuk pengumpulan tugas.
- ☐ Karena tidak diizinkan: Tidak menggunakan bantuan AI dengan cara apa pun dalam pembuatan tugas.

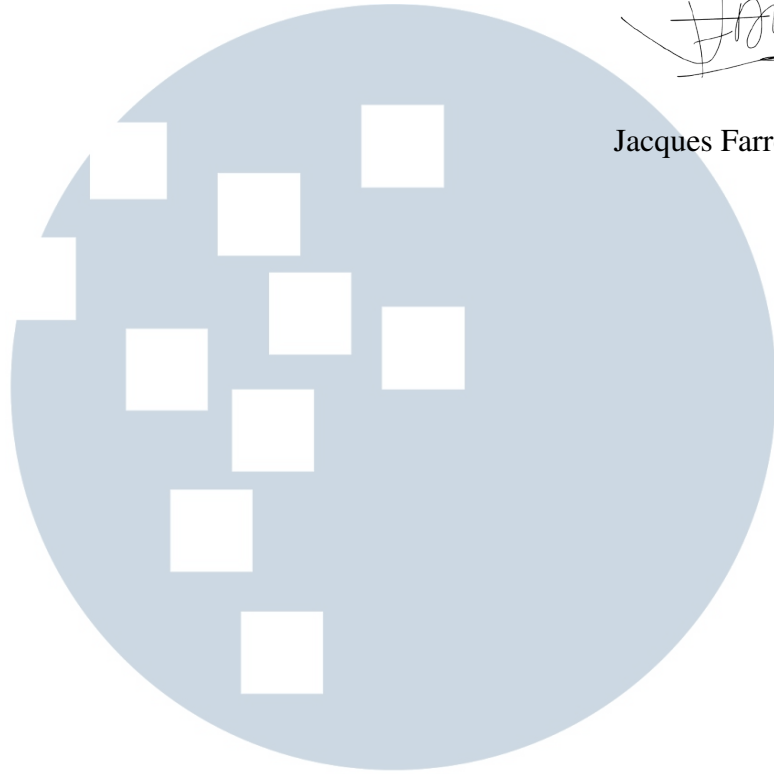
Saya juga menyatakan bahwa:

1. Menyerahkan secara lengkap dan jujur penggunaan perangkat AI yang diperlukan dalam tugas melalui Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI)
2. Mengakui telah menggunakan bantuan AI dalam tugas saya baik dalam bentuk kata, paraphrase, penyertaan ide atau fakta penting yang disarankan oleh AI dan saya telah menyantumkan dalam sitasi serta referensi.
3. Terlepas dari pernyataan di atas, tugas ini sepenuhnya merupakan karya saya sendiri

Tangerang, 20 Desember 2025



Jacques Farrell Dharma



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN

Nama : Jacques Farrell Dharma

NIM : 00000069027

Program Studi : Informatika

Fakultas : Teknik & Informatika

menyatakan bahwa saya melaksanakan kegiatan di:

Nama Perusahaan/Organisasi : RIS

Alamat : Jl. Scientia Boulevard, Gading Serpong, Kel.
Curug Sangereng, Kec. Kelapa Dua, Kab.
Tangerang, Prop. Banten 15810, Indonesia

Email Perusahaan/Organisasi : fti@umn.ac.id

1. Perusahaan/Organisasi tempat saya melakukan kegiatan dapat divalidasi kebenarannya.
2. Jika dikemudian hari, terbukti ditemukan kecurangan/penyimpangan data yang tidak valid di perusahaan/organisasi tempat saya melakukan kegiatan, maka:
 - a. Saya bersedia menerima konsekuensi dinyatakan **TIDAK LULUS** untuk mata kuliah yang telah saya tempuh.
 - b. Saya bersedia menerima semua sanksi yang berlaku sebagaimana ditetapkan dalam peraturan yang berlaku di Universitas Multimedia Nusantara.

Pernyataan ini saya buat dengan sebenar-benarnya dan digunakan sebagaimana mestinya.

Tangerang, 20 Desember 2025

Menyatakan



Eunike Endariahna Surbakti

Menyatakan



Jacques Farrell Dharma

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH MAHASISWA

Yang bertanda tangan di bawah ini:

Nama : Jacques Farrell Dharma
NIM : 00000069027
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Laporan Research & Technology Program

Menyatakan dengan sesungguhnya bahwa:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- ☐ Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 20 Desember 2025



Jacques Farrell Dharma

U M N
U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"I can do all things through Him who strengthens me."

Philippians 4:13 (NASB)



KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala berkat, rahmat, dan penyertaan-Nya sehingga penulis dapat menyelesaikan Laporan Penelitian ini dengan baik. Penyusunan laporan ini dilakukan sebagai salah satu bentuk pelaksanaan kegiatan MBKM sekaligus sebagai sarana untuk menerapkan pengetahuan yang telah diperoleh selama masa studi, khususnya dalam bidang rekayasa perangkat lunak dan pengembangan sistem berbasis kecerdasan buatan.

Penulis menyampaikan ucapan terima kasih kepada:

1. Bapak Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Bapak Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Ibu Eunike Endariahna Surbakti, S.Kom., M.T.I., selaku Pembimbing pertama yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan, serta motivasi selama proses pengerjaan laporan ini.
5. Bapak Yustinus Widya Wiratama, S.Kom., M.Sc., selaku Pembimbing kedua yang telah memberikan masukan dan dukungan yang sangat berarti hingga selesainya laporan penelitian ini.

Penulis menyadari adanya keterbatasan dalam laporan ini dan berharap masukan yang membangun untuk pengembangan penelitian selanjutnya. Semoga karya ini bermanfaat bagi pembaca, berkontribusi pada pengembangan sistem berbasis kecerdasan buatan di bidang kesehatan, serta menambah pengetahuan bagi pihak yang tertarik melanjutkan penelitian serupa.

Tangerang, 20 Desember 2025



Jacques Farrell Dharma

IMPLEMENTASI ARSITEKTUR BACKEND FASTAPI DAN API GATEWAY UNTUK INTEGRASI MULTIPLE MODEL MACHINE LEARNING

Jacques Farrell Dharma

ABSTRAK

Sistem AIRA (AI Research for Oncology) dikembangkan sebagai platform pendukung diagnosis kanker berbasis kecerdasan buatan yang mampu memproses data genomik melalui mekanisme *machine learning inference*. Penelitian ini bertujuan merancang arsitektur *backend* yang efisien, modular, serta mampu menangani beban komputasi tinggi melalui pemisahan peran antara *Cancer Gateway* dan *AI Backend* menggunakan pola *API gateway* dan *containerization Docker*. Metode penelitian meliputi perancangan arsitektur, implementasi layanan *RESTful API*, serta pengujian performa menggunakan Apache JMeter melalui tiga skenario, yaitu pengujian seluruh endpoint API, *load* dan *stress testing* pada *endpoint* prediksi, serta perbandingan performa beberapa model kanker prostat dan payudara berdasarkan variasi fitur seperti GENE, METHYL, dan MIRNA. Hasil pengujian menunjukkan bahwa sistem mampu mempertahankan waktu respons yang stabil hingga sekitar 100 *concurrent users*, dengan rata-rata waktu respons pada model kanker prostat berada pada kisaran 50–70 ms, sedangkan model kanker payudara memiliki latensi yang lebih tinggi, yaitu sekitar 480–3600 ms, bergantung pada kompleksitas model yang digunakan. Pada skenario *stress*, batas kapasitas sistem mulai terlihat pada rentang 100–200 pengguna, dan model dengan kompleksitas tinggi seperti GENE-18, GENE-25, GENE-35, dan GENE-41 menghasilkan waktu respons yang paling besar dibandingkan model berbasis METHYL dan MIRNA. Secara keseluruhan, penelitian ini menunjukkan bahwa performa sistem sangat dipengaruhi oleh kompleksitas model ML yang dijalankan, serta bahwa arsitektur AIRA memberikan dasar yang kuat untuk pengembangan sistem diagnosis berbasis AI yang andal dan terukur.

Kata kunci: AI Backend, Arsitektur sistem, Inference kanker, Load testing, Machine learning

IMPLEMENTATION OF BACKEND ARCHITECTURE USING FASTAPI AND API GATEWAY FOR INTEGRATION OF MULTIPLE MACHINE LEARNING MODELS

Jacques Farrell Dharma

ABSTRACT

The AIRA (AI Research for Oncology) system was developed as an artificial intelligence–driven diagnostic platform designed to process genomic data through machine learning inference. This research aims to design a backend architecture that is efficient, modular, and capable of handling high computational workloads by separating responsibilities between the Cancer Gateway and the AI Backend using an API gateway pattern and Docker-based containerization. The methodology involves system architecture design, the implementation of RESTful API services, and performance evaluation using Apache JMeter across three scenarios: testing all API endpoints, conducting load and stress testing on the prediction endpoint, and comparing the performance of multiple prostate and breast cancer models based on various feature types such as GENE, METHYL, and MIRNA. The results demonstrate that the system maintains stable response times with up to approximately 100 concurrent users, with prostate cancer models achieving average response times in the range of 50–70 ms, while breast cancer models exhibit higher latency, ranging from approximately 480 to 3600 ms depending on model complexity. Under stress conditions, the system’s capacity limits begin to appear in the range of 100–200 concurrent users, and models with higher computational complexity—such as GENE-18, GENE-25, GENE-35, and GENE-41 produce the highest response times relative to METHYL and MIRNA-based models. Overall, the study shows that system performance is strongly influenced by the computational complexity of the machine learning models, and that the AIRA architecture provides a solid foundation for developing scalable and reliable AI-assisted diagnostic systems.

Keywords: AI Backend, Cancer inference, Load testing, Machine learning, System architecture

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN ORISINALITAS	ii
HALAMAN PERNYATAAN PENGGUNAAN BANTUAN KECERDASAN ARTIFISIAL (AI)	iii
HALAMAN PERNYATAAN KEABSAHAN PERUSAHAAN	v
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
HALAMAN PERSEMBAHAN/MOTO	vii
KATA PENGANTAR	viii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
DAFTAR KODE	xv
DAFTAR LAMPIRAN	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	3
1.5 Urgensi Penelitian	3
1.6 Manfaat Penelitian	4
1.7 Luaran Penelitian	4
BAB 2 TINJAUAN PUSTAKA	6
2.1 Arsitektur Backend untuk ML Inference	6
2.1.1 Layered Architecture	6
2.1.2 Cancer Gateway	6
2.1.3 AI Backend	7
2.1.4 Pemisahan Komponen	7
2.1.5 Hybrid Architecture dengan Docker	7
2.2 Machine Learning untuk Deteksi Kanker	8
2.3 API Gateway Pattern	8
2.4 Performance Metrics dan Benchmarking	9
2.4.1 Metrik Performa untuk Backend Systems	9
2.4.2 Metodologi Load Testing	10
2.4.3 Benchmarking ML Inference Performance	10
2.5 Penelitian Terkait	10
2.5.1 Performance Evaluation untuk Backend Systems	11
2.5.2 ML Inference dan Healthcare Systems	11
2.5.3 Research Gap dan Positioning	11
BAB 3 METODE PENELITIAN	12
3.1 Metode Penelitian	12
3.1.1 Model Software Development Life Cycle	14
3.2 Perancangan Arsitektur Backend	15
3.2.1 Layered Architecture Design	15
3.2.2 Deployment Strategy	16
3.2.3 Data Flow Diagram	17
3.2.4 Database Schema	19

3.3	List Service Development	21
3.4	Skenario Load Test dan Stress Testing	22
3.4.1	Lingkungan Pengujian	22
3.4.2	Konfigurasi dan Prosedur Pengujian Menggunakan Apache JMeter	22
3.4.3	Skenario Pengujian Seluruh API	24
3.4.4	Skenario Load Test dan Stress Test untuk Endpoint Prediksi	24
3.4.5	Skenario Perbandingan Model Prostate dan Breast	26
3.4.6	Persiapan Data Uji dan Analisis	27
BAB 4	HASIL DAN PEMBAHASAN	28
4.1	Gambaran Umum	28
4.2	List API Production	28
4.3	Hasil Implementasi	29
4.3.1	Halaman Diagnosis	29
4.3.2	Halaman Upload Data	30
4.3.3	Halaman Hasil Prediksi	30
4.3.4	Integrasi Sistem antara Cancer Gateway dan AI Backend	31
4.4	Hasil Testing	32
4.4.1	Skenario 1 – Pengujian Semua Endpoint (10 dan 100 Users)	32
4.4.2	Skenario 2 – Load Test dan Stress Testing (Endpoint Prediksi)	37
4.4.3	Skenario 3 – Perbandingan Performa Antar Model	45
BAB 5	SIMPULAN SARAN	51
5.1	Simpulan	51
5.2	Saran	52
DAFTAR PUSTAKA		53



DAFTAR TABEL

Tabel 3.1	Daftar Layanan Backend (<i>Service Endpoints</i>)	21
Tabel 3.2	Skenario Pengujian Seluruh API (GET dan POST)	24
Tabel 3.3	Skenario Load Test dan Stress Test Endpoint Prediksi Model Prostate	25
Tabel 3.4	Skenario Load Test dan Stress Test Endpoint Prediksi Model Breast	26
Tabel 3.5	Skenario Perbandingan Model Prostate dan Breast	27
Tabel 4.1	Daftar API Production Sistem AIRA	28
Tabel 4.2	Hasil Pengujian Semua Endpoint – 10 Users	35
Tabel 4.3	Hasil Pengujian Semua Endpoint – 100 Users	36
Tabel 4.4	Hasil Load Test Endpoint Prediksi – Prostate	41
Tabel 4.5	Hasil Load Test Endpoint Prediksi – Breast	41
Tabel 4.6	Hasil Stress Test Endpoint Prediksi – Prostate	42
Tabel 4.7	Hasil Stress Test Endpoint Prediksi – Breast	43
Tabel 4.8	Perbandingan Performa Model Prostate dan Breast	43
Tabel 4.9	Ringkasan Hasil Pengujian Per <i>Model</i> – Prostate (100 pengguna)	48
Tabel 4.10	Ringkasan Hasil Pengujian Per <i>Model</i> – Breast (100 pengguna)	48
Tabel 4.11	Perbandingan Agregat Performa Kelompok <i>Model</i> Prostate dan Breast	49



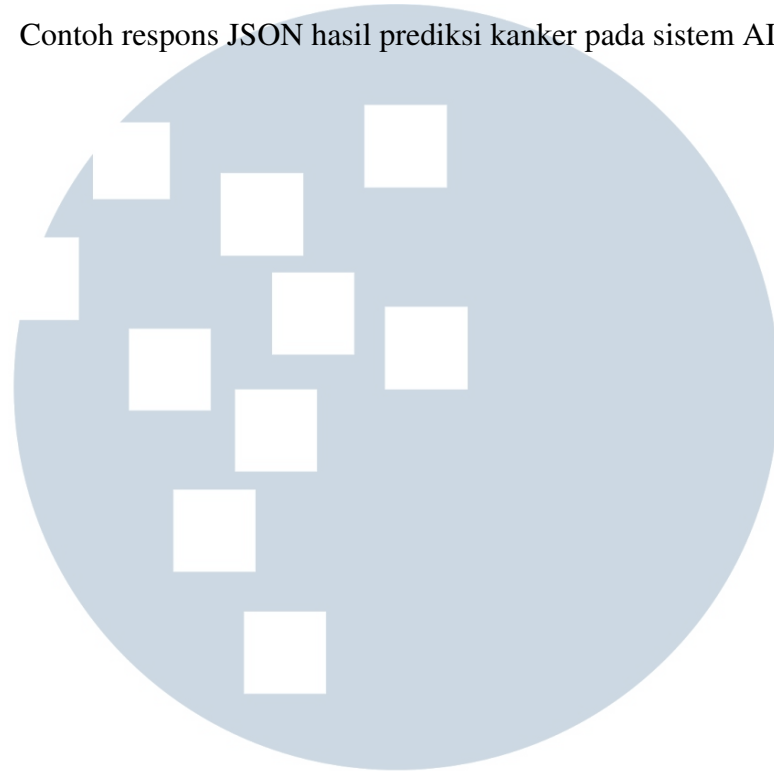
DAFTAR GAMBAR

Gambar 3.1	Research Pipeline Penelitian	12
Gambar 3.2	Layered Architecture Design of AIRA	16
Gambar 3.3	Deployment Architecture Sistem AIRA dengan Docker . .	17
Gambar 3.4	Data Flow Diagram Level 0	18
Gambar 3.5	Data Flow Diagram Level 1	18
Gambar 3.6	AIRA Database Schema	19
Gambar 3.7	Konfigurasi Thread Group pada Apache JMeter untuk Simulasi Beban Pengguna	23
Gambar 3.8	Konfigurasi HTTP Request Sampler untuk Endpoint Prediksi Kanker	23
Gambar 4.1	Halaman Diagnosis – Pemilihan Parameter Model	29
Gambar 4.2	Halaman Unggah CSV untuk Proses Prediksi	30
Gambar 4.3	Halaman Hasil Prediksi Kanker	31
Gambar 4.4	Perbandingan Average Response Time (ART) pada 10 dan 100 pengguna	33
Gambar 4.5	Throughput pada 10 dan 100 pengguna	34
Gambar 4.6	Average Response Time – Prostate	37
Gambar 4.7	Average Response Time – Breast	38
Gambar 4.8	Throughput Comparison – Prostate vs Breast	39
Gambar 4.9	Error Rate – Prostate vs Breast	40
Gambar 4.10	Perbandingan <i>Average Response Time</i> per <i>model</i> (100 pengguna)	46
Gambar 4.11	<i>Error Rate</i> per <i>model</i> pada 100 pengguna	47



DAFTAR KODE

Kode 4.1 Contoh respons JSON hasil prediksi kanker pada sistem AIRA . 32



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

DAFTAR LAMPIRAN

Lampiran 1	PROSTEP-01 Cover Letter MBKM Research	55
Lampiran 2	PROSTEP-02 MBKM Research Card	56
Lampiran 3	PROSTEP-03 Daily Task - MBKM Research	57
Lampiran 4	PROSTEP-04 Verification Form of Research Report MBKM Research	79
Lampiran 5	PROSTEP-05 Letter of Acceptance	80
Lampiran 6	Form Bimbingan	81
Lampiran 7	Hasil Pengecekan Similarity Turnitin	82
Lampiran 8	Formulir Penggunaan Perangkat Kecerdasan Artifisial (AI) . .	89
Lampiran 9	Research Paper	93

