

DAFTAR ISI

LEMBAR PERSEMBAHAN	i
ABSTRAK	iii
ABSTRACT	iv
KATA PENGANTAR.....	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	x
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan.....	4
1.4 Manfaat	4
1.4.1 Manfaat Teoritis	4
1.4.2 Manfaat Praktis	5
1.5 Batasan Masalah	5
1.6 Kerangka Pemikiran	6
1.7 Sistematika Penulisan	9
BAB II TINJAUAN PUSTAKA.....	11
2.1 Hasil Penelitian Terdahulu	11
2.2 Dasar Teori	15
2.2.1 Cloud Computing.....	15
2.2.2 Google Cloud Platform (GCP).....	16
2.2.3 Google Kubernetes Engine (GKE)	16
2.2.4 Sistem Terdistribusi	17
2.2.5 Arsitektur Microservices dan Container Orchestration.....	18
2.2.6 Benchmarking Sistem	18
2.2.7 Workload Artificial Intelligence	19
2.2.8 MobileNetV2	20
2.2.9 Containerized AI Model Serving	20
2.2.10 Autoscaling pada GKE.....	21
2.2.11 Load Balancing pada GKE.....	22
2.2.12 Metrics Evaluasi Kinerja Cloud	23
2.2.13 Diminishing Returns dalam Scaling.....	23

2.2.14 Desain Eksperimen dan Validasi Statistik	24
2.2.15 Optimasi Kinerja Inferensi AI di Cloud.....	27
BAB III METODOLOGI PENELITIAN	28
3.1 Inisiasi.....	28
3.1.1 Identifikasi Masalah	28
3.1.2 Perumusan Hipotesis.....	29
3.1.3 Perancangan Eksperimen	30
3.2 Pengumpulan Data.....	33
3.2.1 Konfigurasi Infrastruktur	33
3.2.2 Deployment Model AI MobileNetV2 & Service Load Balancer.....	35
3.2.3 Eksekusi Load Testing	36
3.2.4 Ekstraksi Data	37
3.3 Analisis Data.....	37
3.3.1 Validasi Kebersihan Data.....	37
3.3.2 Uji Inferensial.....	38
3.4 Benchmark Report.....	39
3.5 Action Plan	39
BAB IV HASIL DAN PEMBAHASAN	40
4.1 Hasil Penelitian.....	40
4.1.1 Inisiasi	40
4.1.2 Pengumpulan Data	42
4.1.3 Analisis Data	44
4.1.4 Benchmark Report	50
4.1.5 Action Plan.....	58
4.2 Pembahasan	58
4.2.1 Rumusan Masalah 1	58
4.2.2 Rumusan Masalah 2	60
4.2.3 Validasi Kuantitatif dan Skema Solusi Terapan	62
BAB V SIMPULAN DAN SARAN	65
5.1 Kesimpulan.....	65
5.2 Action Plan	66
5.3 Saran	66
DAFTAR PUSTAKA	68
DAFTAR LAMPIRAN	72

DAFTAR GAMBAR

Gambar 1.1 Diagram Alur Kerangka Pemikiran Penelitian	7
Gambar 2.1 Tabel Prisma	11
Gambar 3.1 Diagram Alur Penelitian	28
Gambar 3.2 Tahap Inisiasi.....	28
Gambar 3.3 System Under Test.....	31
Gambar 3.4 Pengumpulan Data	33
Gambar 3.5 Topologi Arsitektur Jaringan dan Skema Skenario	35
Gambar 3.6 Analisis Data.....	37
Gambar 4.1 Grafik RPS dan Average Latency (Skenario A)	53
Gambar 4.2 Grafik RPS dan Average Latency (Skenario B)	53
Gambar 4.3 Grafik RPS dan Average Latency Lintas (Skenario C)	54
Gambar 4.4 Error Rate (%) Lintas Skenario dan Beban Kerja	55
Gambar 4.5 Grafik Komparasi Metrik Layer Kernel Skenario C	56



DAFTAR TABEL

Tabel 2.1 State of the Arts	12
Tabel 2.2 Matriks Rumus Kalkulasi Standar Analisis Two-Way ANOVA	27
Tabel 3.1 Spesifikasi Perangkat Lunak	34
Tabel 4.1 Matriks Koefisien Variasi (CV) Parameter Kunci Seluruh Skenario... 44	
Tabel 4.2 Hasil Uji Normalitas Shapiro-Wilk pada Residual Model.....	46
Tabel 4.3 Hasil Uji Homogenitas Varians Levene's Test	47
Tabel 4.4 Hasil Analisis Varians Faktorial (Two-Way ANOVA)	49
Tabel 4.5 Matriks Komparasi Kinerja Sisi Klien (Locust Metrics)	50
Tabel 4.6 Matriks Utilitas Sumber Daya Klaster GKE	51
Tabel 4.7 Rekapitulasi Log Forensik Siklus Hidup Pod Skenario C	57

