



BERITA ACARA SIDANG PENDADARAN TUGAS AKHIR

S/UBL/FTI/0897/VII/26

Pada hari ini, Kamis 16 Juli 2026 telah dilaksanakan Ujian Sidang Pendadaran Tugas Akhir sebagai berikut:

Judul: PEMBANGUNAN SISTEM PREDIKSI SATURASI STORAGE ELASTICSEARCH
MENGUNAKAN MULTIPLE LINEAR REGRESSION DAN DECISION TREE
REGRESSION PADA PT EBDESK TEKNOLOGI

Nama : Lingga Andika Setiabudi
NIM : 2211510637
Dosen Pembimbing : Joko Christian Chandra, S.Kom., M.Kom.

Berdasarkan penilaian pada Presentasi + Demo, Penulisan, Penguasaan Materi, Penguasaan Program maka Mahasiswa tersebut di atas dinyatakan:

LULUS

dengan nilai angka : **89** huruf : **A**

Mahasiswa tersebut di atas wajib menyerahkan hasil perbaikan tulisan Tugas Akhir dalam bentuk terjilid sesuai dengan Panduan Perbaikan Tugas Akhir, selambat-lambatnya Kamis 30 Juli 2026.

Panitia Penguji:

- 1 Ketua Ir. Gunawan Pria Utama, S.Kom., M.Kom.
- 2 Anggota Prof. Dr. Ir. Arief Wibowo, M.Kom.
- 3 Moderator Joko Christian Chandra, S.Kom., M.Kom.

Keterangan:

Nilai Huruf: A:85-100 A-:80-84,99 B+:75-79,99 B:70-74,99 B-:65-69,99 C:60-64,99 D:40-59,99
E-:0-39,99



UNIVERSITAS BUDI LUHUR

FAKULTAS TEKNOLOGI INFORMASI

Kartu Bimbingan Tugas Akhir

NIM: 2211510637

Nama: Lingga Andika Setiabudi

Pembimbing: Joko Christian Chandra, S.Kom., M.Kom.

No.	Tanggal	Materi
1	16-04-2026	diskusi tema terkait implementasi multivariate regresion algorithm untuk prediksi dan warning kapasista penyimpanan client di data center
2	29-04-2026	draft bab 1, arahan penulisan rumusan masalah dan batasan masalah
3	11-05-2026	perbaikan bab 1, arahan penentuan metodologi riset pada bab 2 dan draft bab 2
4	14-05-2026	perbaikan bab 2 dan arahan penulisan bab 3 perbaikan
5	29-05-2026	Perbaikan bab 3, penggunaan gambar dan penjelasan gambar
6	09-06-2026	Perbaikan bab 3, demo output machine learning
7	25-06-2026	demo perbaikan, draft bab 4 dan arahan perbaikan bab 4
8	02-07-2026	perbaikan bab 4 dan bab 5



LEMBAR PENGESAHAN

Nama : Lingga Andika Setiabudi
Nomor Induk Mahasiswa : 2211510637
Program Studi : Teknik Informatika
Bidang Peminatan : Programming Expert
Jenjang Studi : Strata 1
Judul : PEMBANGUNAN SISTEM PREDIKSI SATURASI STORAGE
ELASTICSEARCH MENGGUNAKAN MULTIPLE LINEAR
REGRESSION DAN DECISION TREE REGRESSION PADA PT EBDESK
TEKNOLOGI



Laporan Tugas Akhir ini telah disetujui, disahkan dan direkam secara elektronik sehingga tidak memerlukan tanda tangan tim penguji.

Jakarta, Kamis 16 Juli 2026

Tim Penguji:

Ketua : Gunawan Pria Utama, S.Kom., M.Kom.
Anggota : Prof. Dr. Ir. Arief Wibowo, M.Kom.
Pembimbing : Joko Christian Chandra, S.Kom., M.Kom.
Ketua Program Studi : Dr. Indra, S.Kom., M.T.I.

**PEMBANGUNAN SISTEM PREDIKSI SATURASI STORAGE
ELASTICSEARCH MENGGUNAKAN MULTIPLE LINEAR REGRESSION
DAN DECISION TREE REGRESSION PADA PT EBDESK TEKNOLOGI**

TUGAS AKHIR



Oleh:

**LINGGA ANDIKA SETIABUDI
NIM: 2211510637**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

**JAKARTA
2025/2026**

ABSTRAK

PEMBANGUNAN SISTEM PREDIKSI SATURASI STORAGE ELASTICSEARCH MENGGUNAKAN MULTIPLE LINEAR REGRESSION DAN DECISION TREE REGRESSION PADA PT EBDESK TEKNOLOGI

Oleh: Lingga Andika Setiabudi (2211510637)

Pertumbuhan *storage* pada klaster Elasticsearch di PT eBdesk Teknologi bersifat monoton tanpa sistem prediksi kapasitas yang proaktif, sehingga penelitian ini bertujuan membangun sistem prediksi saturasi storage yang memberikan peringatan dini bagi tim operasional. Sistem dibangun dengan pendekatan regresi *autoregresif* berbasis riwayat storage dan pola temporal, dengan telemetri *throughput* Kafka diuji kontribusinya sebagai fitur tambahan, menggunakan dua algoritma *from-scratch*: Multiple Linear Regression dengan regularisasi Ridge (MLR+Ridge) dan Decision Tree Regression (DTR). Data dikumpulkan dari *Prometheus* selama 16 hari (20 Juni–5 Juli 2026, interval lima menit), menghasilkan 4.608 baris dan 27 fitur, dengan horizon prediksi 24 jam sebagai horizon utama dan 7 hari sebagai eksplorasi awal. Model MLR+Ridge formulasi delta tanpa fitur Kafka mencapai Mean Absolute Error (MAE) 25,7 GB pada horizon 24 jam, praktis setara DTR (25,9 GB), ditetapkan sebagai model operasional berdasarkan pertimbangan multi-kriteria (akurasi, kemampuan ekstrapolasi, interpretabilitas), dan mengungguli *baseline persistence* 11,0 kali lipat. Formulasi delta terbukti esensial mengatasi sifat non-stasioner data, sementara analisis ablasi dan *feature importance* menunjukkan kontribusi Kafka marjinal dan tidak konsisten ($<0,1\%$), mengindikasikan riwayat storage sudah menyerap informasi tersebut secara implisit. Berdasarkan laju pertumbuhan 0,315% per hari dari utilisasi 76,9%, sistem mengestimasi saturasi dalam ≈ 26 hari (sekitar 31 Juli 2026) meskipun estimasi ini bersifat ekstrapolasi linear, bukan prediksi model yang tervalidasi langsung untuk horizon sejauh itu. Sistem diintegrasikan sebagai antarmuka Streamlit dengan *endpoint* /metrics yang kompatibel dengan Prometheus untuk pemantauan berkelanjutan.

Kata Kunci: Elasticsearch, prediksi storage, *autoregressive forecasting*, capacity planning, Multiple Linear Regression, Decision Tree Regression, *time series forecasting*.

DAFTAR ISI

LEMBAR PENGESAHAN	iii
ABSTRAK.....	iv
SURAT PERNYATAAN TIDAK PLAGIAT	v
SURAT PERNYATAAN PENGGUNAAN AI GENERATIF	vi
KATA PENGANTAR	vii
DAFTAR GAMBAR.....	viii
DAFTAR TABEL	x
DAFTAR SIMBOL	xi
DAFTAR ALGORITMA	xii
DAFTAR ISI	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah.....	4
1.3 Batasan Masalah.....	4
1.4 Tujuan dan Manfaat.....	6
1.4.1 Tujuan.....	6
1.4.2 Manfaat.....	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI.....	8
2.1 Data Mining.....	8
2.2 Knowledge Discovery in Databases (KDD).....	8
2.3 Apache Kafka	10
2.3.1 Arsitektur Apache Kafka.....	10
2.3.2 Kafka Throughput Metrics	10
2.3.3 Mekanisme Fan-Out pada Pipeline Percolation	11
2.4 Elasticsearch.....	11
2.4.1 Elasticsearch Filesystem Metrics.....	12
2.4.2 Kondisi Saturasi Storage	12
2.5 Integrasi Apache Kafka dan Elasticsearch	13
2.5.1 Pola Arsitektur Kafka sebagai Upstream Elasticsearch.....	13
2.5.2 Consumer Lag sebagai Fitur Operasional	14
2.5.3 Fan-Out sebagai Pengali Non-Linear	14
2.5.4 Arsitektur Integrasi pada Penelitian Ini	15
2.6 Manajemen Kapasitas Proaktif.....	16
2.7 Peramalan Deret Waktu dan Model Autoregresif	16
2.7.1 Strategi Prediksi Direct Multi-Horizon	18
2.8 Multiple Linear Regression dengan Ridge Regularization.....	19
2.8.1 Persamaan Model	19
2.8.2 Ridge Regularization (L2).....	19
2.8.3 Gradient Descent dan Normalisasi Fitur	20
2.9 Decision Tree Regression.....	20
2.9.1 Kriteria Split	20
2.9.2 Stopping Criteria dan Prediksi.....	21
2.9.3 Feature Importance.....	21
2.10 Metrik Evaluasi Model.....	22

2.10.1 Mean Absolute Error (MAE).....	22
2.10.2 Root Mean Squared Error (RMSE)	22
2.10.3 Mean Absolute Percentage Error (MAPE)	23
2.11 Penelitian Terkait.....	23
BAB III METODOLOGI PENELITIAN	30
3.1 Data Penelitian.....	30
3.1.1 Stack Monitoring.....	30
3.1.2 Pipeline Sumber Fitur Prediktor	31
3.1.3 Skema Dataset	31
3.1.4 Arsitektur Sistem yang Dibangun.....	32
3.2 Metode Pembandingan	33
3.3 Metode Penelitian.....	36
3.3.1 Seleksi Data	37
3.3.2 Pra-pemrosesan.....	40
3.3.3 Transformasi.....	41
3.3.4 Pemodelan	43
3.3.5 Interpretasi dan Visualisasi.....	49
3.3.6 Evaluasi	51
3.4 Rancangan Pengujian	53
3.4.1 Pengujian Akurasi Model Prediksi	53
3.4.2 Pengujian Feature Importance dan Koefisien.....	54
3.4.3 Pengujian Fungsionalitas Sistem	55
3.5 Rancangan Menu	56
3.6 Rancangan Layar	58
3.6.1 Rancangan Halaman Sumber Data	58
3.6.2 Rancangan Halaman Pra-pemrosesan	59
3.6.3 Rancangan Halaman Transformasi.....	60
3.6.4 Rancangan Halaman Pemodelan	61
3.6.5 Rancangan Halaman Evaluasi	62
3.6.6 Rancangan Halaman Prediksi dan Output	63
3.6.7 Rancangan Dashboard Grafana	64
BAB IV HASIL DAN PEMBAHASAN	66
4.1 Lingkungan Percobaan	66
4.1.1 Spesifikasi Perangkat Keras	66
4.1.2 Spesifikasi Perangkat Lunak	67
4.1.3 Deployment Diagram	68
4.2 Implementasi Metode	69
4.2.1 Seleksi Sumber Fitur Prediktor.....	69
4.2.2 Karakteristik Dataset	70
4.2.3 Pra-pemrosesan.....	72
4.2.4 Transformasi dan Rekayasa Fitur	73
4.3 Pemodelan	78
4.3.1 Formulasi Target: Absolut versus Delta	78
4.3.2 MLR dengan Ridge Regularization.....	79
4.3.3 Decision Tree Regression.....	80
4.3.4 Interpretasi dan Visualisasi.....	80
4.4 Flowchart	81

4.4.1 Flowchart Pipeline Metode End-to-End.....	81
4.4.2 Flowchart Algoritma MLR dengan Ridge Regularization	82
4.4.3 Flowchart Algoritma DTR.....	83
4.4.4 Flowchart Tahap Pra-pemrosesan	84
4.4.5 Flowchart Tahap Transformasi.....	85
4.5 Algoritma.....	86
4.5.1 Algoritma Pra-pemrosesan	86
4.5.2 Algoritma Transformasi per Horizon	88
4.5.3 Algoritma MLR dengan Ridge Regularization	89
4.5.4 Algoritma Decision Tree Regression	90
4.5.5 Algoritma Evaluasi dan Estimasi Saturasi.....	92
4.6 Hasil dan Evaluasi	93
4.6.1 Akurasi Model: Formulasi Absolut vs Delta (Horizon 24 Jam).....	93
4.6.2 Perbandingan MLR+Ridge vs DTR	95
4.6.3 Ablation Fitur Kafka: Kontribusi Telemetry Hulu.....	96
4.6.4 Feature Importance dan Koefisien Model	98
4.6.5 Proyeksi Saturasi Storage	101
4.6.6 Evaluasi Hasil Horizon 7 Hari.....	103
4.6.7 Ringkasan Hasil Evaluasi	104
4.6.8 Pengujian Fungsionalitas Sistem	106
4.7 Tampilan Layar Antarmuka Sistem.....	107
BAB V PENUTUP	112
5.1 Kesimpulan.....	112
5.2 Saran	113
LAMPIRAN	117

DAFTAR PUSTAKA

- Azure, I., 2024. Predictive modeling for industrial productivity: Evaluating linear regression and decision tree regressor approaches. *Journal of AppliedMath* 2(4), 1435. <https://doi.org/10.59400/jam.v2i4.1435>
- Bisdoulis, K.-L., 2025. Assets forecasting with feature engineering and transformation methods for LightGBM. arXiv:2501.07580. <https://arxiv.org/abs/2501.07580>
- Bontempi, G., Ben Taieb, S., Le Borgne, Y.-A., 2012. Machine learning strategies for time series forecasting. In: Aufaure, M.-A., Zimányi, E. (Eds.), *European Business Intelligence Summer School (eBISS 2012), Lecture Notes in Business Information Processing*, vol. 138. Springer, Berlin, pp. 62–77. https://doi.org/10.1007/978-3-642-36318-4_3
- Calderon, G., del Campo, G., Saavedra, E., Santamaría, A., 2024. Monitoring framework for the performance evaluation of an IoT platform with Elasticsearch and Apache Kafka. *Information Systems Frontiers* 26(6), 2373–2389. <https://doi.org/10.1007/s10796-023-10409-2>
- Elastic, n.d. Disk-based shard allocation. *Elasticsearch Guide*. <https://www.elastic.co/guide/en/elasticsearch/reference/current/disk-allocator.html>
- He, Y.-X., Lyu, S.-H., Jiang, Y., 2025. Interpreting deep forest through feature contribution and MDI feature importance. *ACM Transactions on Knowledge Discovery from Data* 20(1), Article 15, 21 pp. <https://doi.org/10.1145/3641108>
- Hyndman, R.J., Athanasopoulos, G., 2021. *Forecasting: Principles and Practice*, 3rd ed. OTexts, Melbourne. <https://otexts.com/fpp3/>
- Juanta, P., Djuli, Z., Tifanny, Sitanggang, D., Anita, 2025. Comparing regression methods for assessing and prediction the impact of salary increases on employee performance. *JIKO (Jurnal Informatika dan Komputer)* 8(3), 173–180. <https://doi.org/10.33387/jiko.v8i3.10098>
- Kirchoff, D.F., Meyer, V., Calheiros, R.N., De Rose, C.A.F., 2024. Evaluating machine learning prediction techniques and their impact on proactive resource provisioning for cloud environments. *The Journal of Supercomputing* 80(15), 21920–21951. <https://doi.org/10.1007/s11227-024-06303-6>
- Kong, F., Chua, F.-F., Lim, A.H.-L., 2026. Resource cost management in cloud service environment. *Bulletin of Electrical Engineering and Informatics* 15(1), 470–483. <https://doi.org/10.11591/eei.v15i1.10352>

- Masich, I.S., Tynchenko, V.S., Nelyub, V.A., Bukhtoyarov, V.V., Kurashkin, S.O., Gantimurov, A.P., Borodulin, A.S., 2022. Prediction of critical filling of a storage area network by machine learning methods. *Electronics* 11(24), 4150. <https://doi.org/10.3390/electronics11244150>
- Noetzold, D., Rossetto, A.G.D.M., Leithardt, V.R.Q., Costa, H.J.M., 2024. Enhancing infrastructure observability: Machine learning for proactive monitoring and anomaly detection. *Journal of Internet Services and Applications* 15(1). <https://doi.org/10.5753/jisa.2024.4509>
- Nurani, A.T., Setiawan, A., Susanto, B., 2023. Perbandingan kinerja regresi Decision Tree dan regresi linear berganda untuk prediksi BMI pada dataset Asthma. *Jurnal Sains dan Edukasi Sains* 6(1), 34–43. <https://doi.org/10.24246/juses.v6i1p34-43>
- Papakyriakou, D., Barbounakis, I.S., 2022. Data mining methods: A review. *International Journal of Computer Applications* 183(48), 5–19. <https://www.ijcaonline.org/archives/volume183/number48/papakyriakou-2022-ijca-922009.pdf>
- Raptis, T.P., Passarella, A., 2023. A survey on networked data streaming with Apache Kafka. *IEEE Access* 11, 85333–85350. <https://doi.org/10.1109/ACCESS.2023.3303810>
- Rawat, G.S., 2025. Towards intelligent data retention recommendations in DevOps using Elasticsearch and ML. *SSRG International Journal of Computer Science and Engineering* 12(8), 1–12. <https://doi.org/10.14445/23488387/IJCSE-V12I8P101>
- Ren, Y., 2025. An empirical analysis of overfitting mitigation techniques using regularized linear regression: A case study on noise data modeling. *Advances in Engineering Technology Research* 14 (Proceedings of BEMSIC 2025), 1095–1100. ISSN: 2790-1688.
- Saory, F., Jaman, J.H., Carudin, 2025. Sistem monitoring server menggunakan Prometheus dan Grafana dengan integrasi Slack di PT. Atlas Lintas Indonesia. *JITET (Jurnal Informatika dan Teknik Elektro Terapan)* 13(3S1), 1150–1157. <https://doi.org/10.23960/jitet.v13i3S1.8024>
- Shabana, L.I., Iqbal, M., 2025. Perancangan sistem monitoring dan forecasting sumber daya server berbasis Grafana di PT. Dirgantara Indonesia. *e-Proceeding of Applied Science* 11(4), 1019–1024. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/appliedscience/article/view/27789>
- Singhal, N., Himanshu, 2022. A review on knowledge discovery from databases. In: *Lecture Notes in Electrical Engineering*, vol. 814. Springer, Singapore, pp. 443–452. https://doi.org/10.1007/978-981-16-9488-2_43

- Vyas, S., Tyagi, R.K., Jain, C., Sahu, S., 2022. Performance evaluation of Apache Kafka – A modern platform for real time data streaming. In: 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM). IEEE. <https://doi.org/10.1109/ICIPTM54933.2022.9754154>
- Wang, Y., Liang, Y., Zhang, B., Yi, S., Zhou, P., Lan, X., Xue, C., Li, Y., Li, X., Xu, C., 2025. Forecasting tuberculosis epidemics using an autoregressive fractionally integrated moving average model: A 17-year time series analysis. *Journal of Global Health* 15, 04208. <https://doi.org/10.7189/jogh.15.04208>
- Yang, Y., Wang, S., Laporte, G., 2025. Improved regression tree models using generalization error-based splitting criteria. *Naval Research Logistics* 72, 1097–1116. <https://doi.org/10.1002/nav.22270>