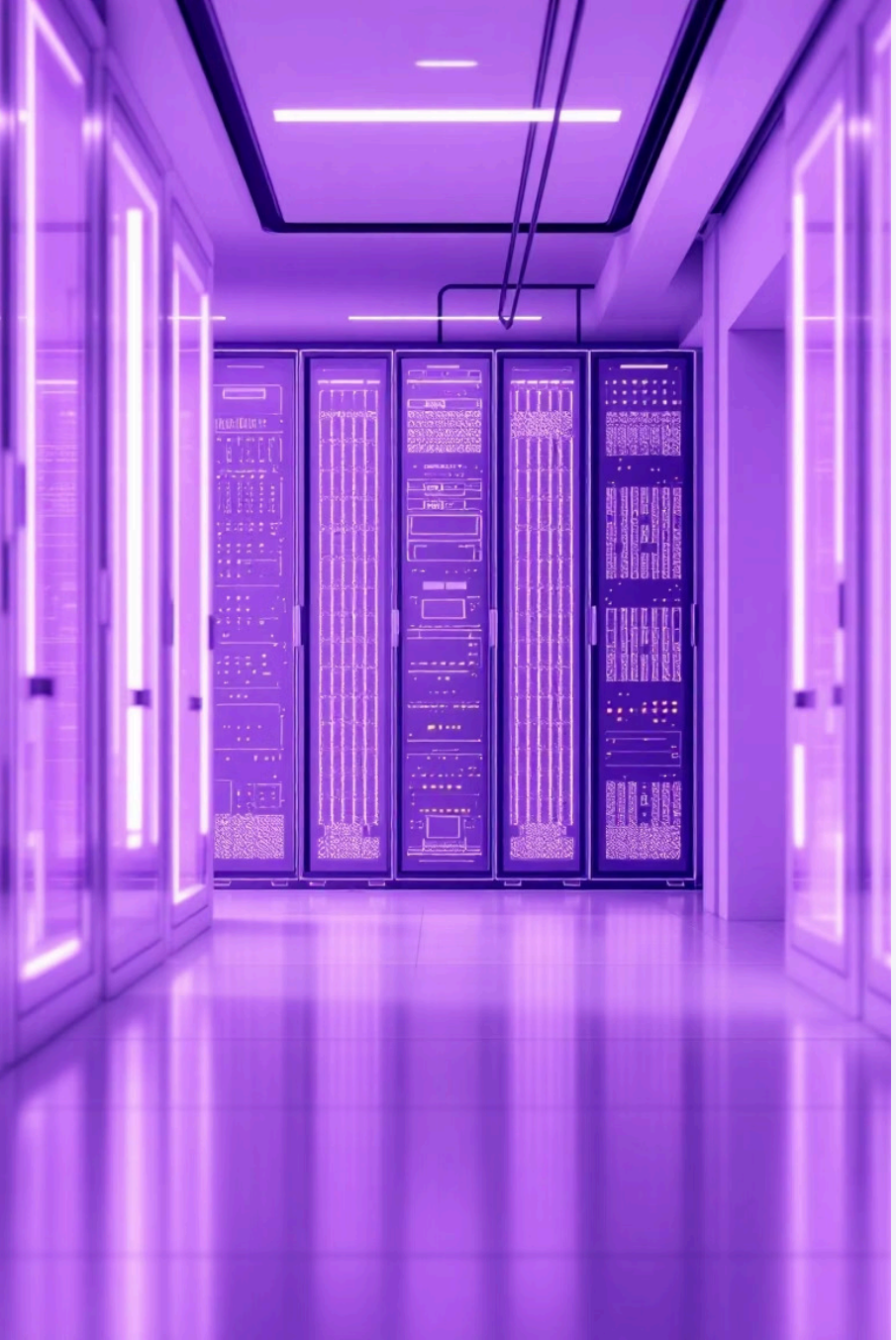


Regression-Based Network Load Forecasting for Sustainable Digital Infrastructure

A pilot project using historical data to predict network traffic patterns, enabling smarter capacity decisions that reduce costs and environmental impact.

Project Author: Aurora A

Tomorrow University – Regression Capstone

Current Bandwidth Decisions: Reactive, Not Strategic

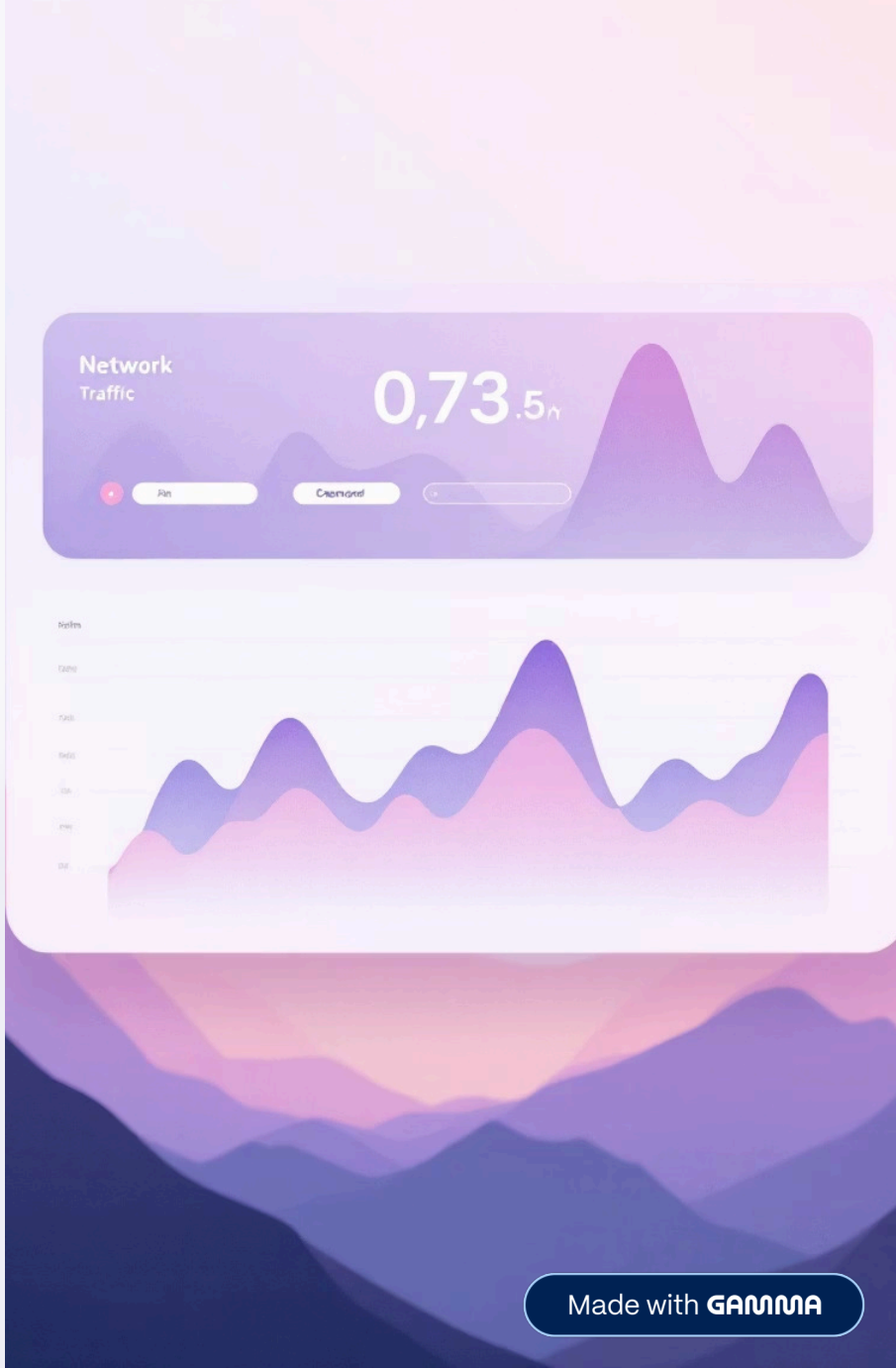
How decisions are made today

Our network teams typically overprovision capacity "just in case" and react to traffic spikes as they occur. This approach lacks visibility into future demand patterns.

Our providers bill based on the 95th percentile of usage, making peak understanding critical for cost control.

Why this creates problems

- **Financial:** Paying for capacity that sits idle most of the time
- **Operational risk:** Underprovisioning leads to congestion and emergency upgrades
- **Sustainability:** Idle hardware consumes power and cooling without benefit



Scope: A Focused Pilot to Test the Concept

What we used

- 1 edge router
- 2 upstream providers (uplinks)
- 3 backend hosts behind the router
- Approximately 5 months of hourly operational data

What we predict

Next-hour peak traffic on the upstream connection. This is not yet full monthly billing metrics, but a realistic foundation for testing forecasting capability

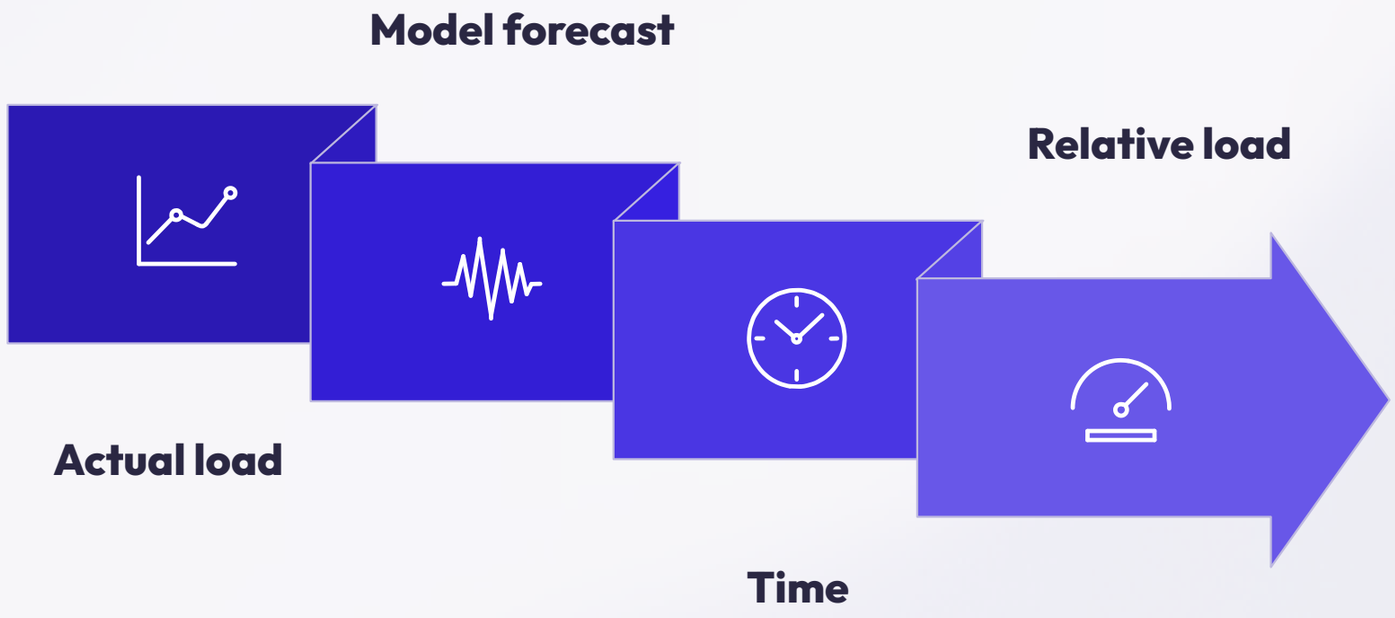
Why this scope

Deliberately narrow pilot designed to answer one question: can we reliably forecast near-term traffic peaks using aggregated operational data?

How the Forecasting Pipeline Works

- Uses only **aggregated hourly operational data** — no user or content-level information required
- Features are **leakage-safe**: model sees only historical information available at prediction time
- Selected **Ridge regression** after testing multiple models for best balance of accuracy, stability, and interpretability

Model Performance: Capturing Traffic Patterns



Illustrative example of how forecasts follow the overall traffic pattern. For real metrics and plots, see the technical notebook

- The model **consistently performs better** than simple baselines like 'next hour = last hour'.
- It captures typical **daily and weekly patterns**, giving a reliable sense of where the load is heading in the short term.
- It still cannot predict every **rare extreme spike exactly**, but usually recognises that a heavy hour is likely.

Practical Applications: Using Forecasts Today

Operational Scheduling

Delay non-urgent, bandwidth-heavy tasks like backups and large synchronisations during hours flagged as likely high load.

Early Warning System

Monitor forecast frequency approaching chosen utilisation thresholds as signal to review tariffs or capacity before issues arise.

Future Tariff Decisions

Use forecast patterns as evidence for upgrade/downgrade discussions, moving beyond intuition to data-driven planning.

This is decision support, not automatic scaling — humans remain in control with conservative safety margins.

Sustainability Benefits: Right-Sizing Infrastructure

Better Forecasts

Accurate traffic predictions reduce uncertainty

Right-Sized Capacity

Match hardware to actual demand patterns

Less Idle Hardware

Fewer ports and devices consuming power unnecessarily

Lower Emissions

Reduced power and cooling needs

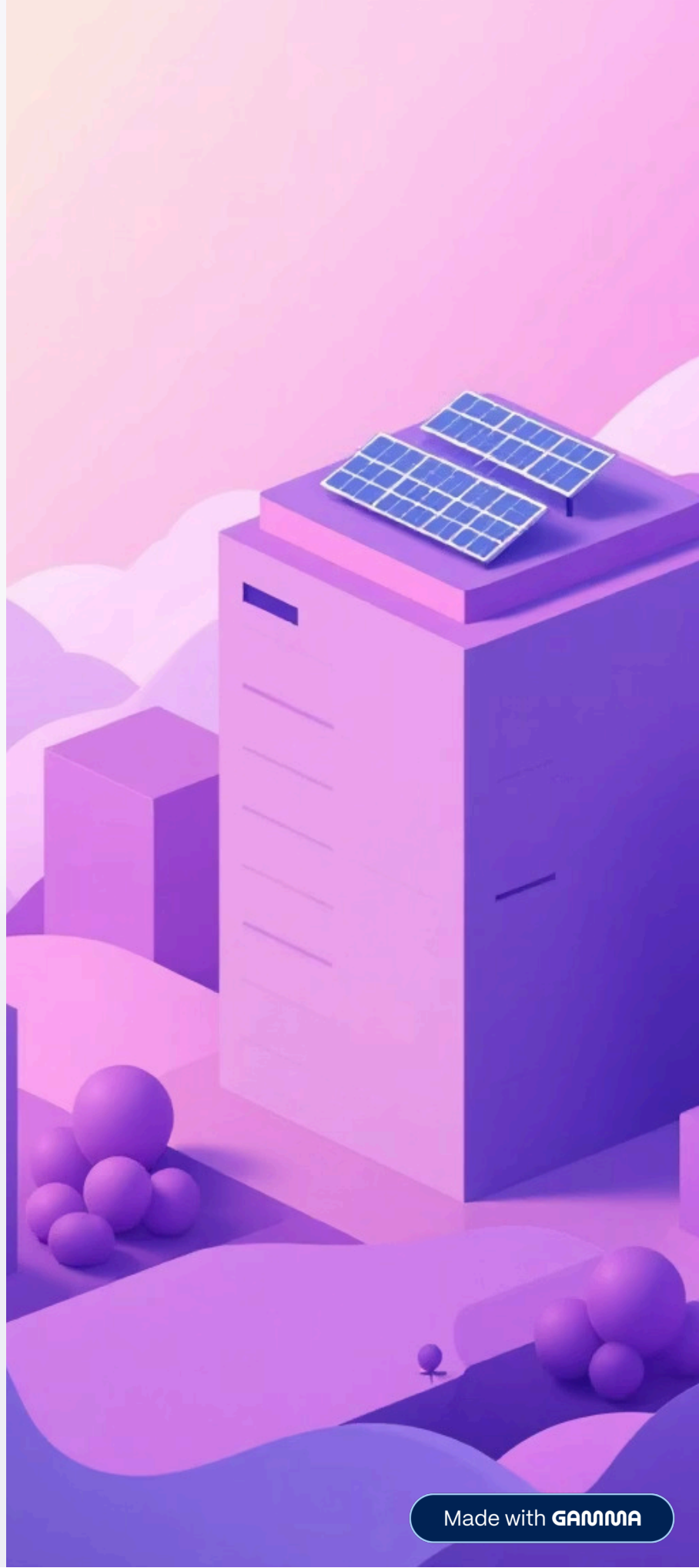
Key sustainability impacts

- Fewer idle ports and devices consuming power and cooling
- Opportunity to shift heavy work to quieter windows
- Alignment with greener hours in energy mix

Important context

Reliability and sustainability treated as **joint constraints** — not sacrificing resilience to chase savings.

Supports **SDG 9** (Industry, Innovation), **SDG 12** (Responsible Consumption), **SDG 13** (Climate Action)



Limitations & Guardrails: Using the Model Safely

Data scope

Approximately 5 months of data from **one router** — does not yet cover other sites or full annual patterns including seasonal variations.

Operating conditions

Works best under **"business as usual"** conditions. Major architectural or product changes require retraining and revalidation.

Forecast nature

Indicates **likelihood and trends, not guarantees**. One signal among others for decision-making.

Guardrails for responsible use

Use as decision-support tool only. Maintain conservative safety margins. Reassess performance regularly as traffic patterns evolve. Do not rely solely on forecasts for critical decisions.

Implementation Roadmap: From Pilot to Production

Phase 1: Proof of Concept

Current state — 1 router, 1-hour forecast horizon, proof that next-hour peaks are predictable enough to be useful for operational decisions.

Phase 2: Expansion

Next step — Extend to additional routers and data centres. Collect ≥12 months of history. Move to 24–168 hour horizons. Begin aggregating into experimental monthly p95 risk indicators.

Phase 3: Integration

Longer term — Integrate forecasts into dashboards for network and sustainability teams. Run pilots where scheduling and tariff discussions informed by model. Continuously refine with new features.

This roadmap shows how to scale safely and incrementally if stakeholders choose to continue investment.

Key Takeaways & Next Steps

What we built

Working forecasting model using historical operational data to predict next-hour network load on one router with demonstrable accuracy.

Why it matters

Enables proactive capacity planning that reduces costs, improves reliability, and lowers environmental impact through right-sized infrastructure.

What's next

Phase 2 expansion requires stakeholder approval, additional data collection, and integration planning. Ready to discuss feasibility and priorities.

Ready for discussion

- Continue to Phase 2 expansion
- Refine use cases with operational teams
- Explore dashboard integration options
- Assess data collection requirements

Decision factors

Business value, resource availability, and strategic alignment with sustainability goals.