

ML LABS INTELLIGENCE

# Global Telecom Roaming Cost Optimization System

A telecom operator was losing margin on international roaming across carrier agreements. ML LABS built the ML routing system that cut cost per session while traffic was live.

CASE STUDY

Author Omar Trejo

Date 2026-02-16

ML LABS

[mlabs.com/intel/global-roaming-optimization](https://mlabs.com/intel/global-roaming-optimization)

---

A top 10 global telecom company was hemorrhaging margin on international roaming — thousands of bilateral carrier agreements across hundreds of markets, each with different pricing tiers, and routing decisions that depended on manual analysis teams couldn't complete before traffic patterns had already shifted. The gap between what the company was paying and what optimized routing could achieve was large enough to justify a dedicated ML platform.

ML LABS was engaged through Gigster to design and build a custom ML platform that could analyze roaming traffic worldwide and optimize both cost and routing decisions using time-series models processing 1-5TB of new data daily.

## The Problem

International roaming involves thousands of bilateral agreements between carriers across hundreds of markets. Each agreement has different pricing tiers, quality-of-service levels, and capacity constraints.

- Routing decisions were based on static rules that didn't adapt to real-time conditions
- Cost optimization required manual analysis that lagged behind traffic changes
- No unified view across markets to identify arbitrage or inefficiency
- Data volume at 1-5TB per day made manual analysis impossible at the required speed

## What ML LABS Built

The engagement delivered a production ML platform with three core capabilities:

1. Time-series models trained on global roaming traffic patterns to forecast demand and cost across markets
2. Cost optimization engine that identified the lowest-cost routing paths while maintaining quality thresholds
3. Routing recommendation system that adapted to shifting traffic patterns and agreement terms in near real time

The platform ingested 1-5TB of new data daily — live network telemetry, agreement terms, and historical traffic patterns — to produce routing decisions that balanced cost, quality, and capacity constraints across the worldwide network.

## Architecture

The system was built as a production platform for operational teams, not a research prototype:

- Streaming data pipeline capable of ingesting and processing terabytes of daily traffic data
- Time-series ML models trained on historical patterns and continuously updated with live data
- Decision layer that produced actionable routing recommendations within operational latency requirements
- Dashboard layer for network operations teams to monitor, override, and audit decisions

## Delivery Pattern

The engagement followed ML LABS' standard execution model: define the highest-value target first, ship a working system fast, then iterate based on production feedback.

- Scoping identified the highest-cost routing corridors as the initial optimization target
- First production deployment focused on a subset of markets to validate the time-series approach
- Expanded coverage worldwide once the models proved reliable under real traffic conditions

## Results

Roaming cost per session decreased measurably across targeted corridors within the first quarter of deployment. The platform identified optimization opportunities across 128% more corridors than the original scope targeted — surfacing routing inefficiencies and cost-saving opportunities that the manual analysis team had never reached. The routing model processed 1-5TB of daily traffic data across global markets — a volume that had made manual optimization physically impossible. Decision latency dropped from manual review cycles measured in days to sub-second automated routing, allowing the network to respond to traffic pattern shifts as they occurred rather than after margin had already been lost. Measured against engagement cost, the routing optimization delivered over 12x return in the first year of operation.

The platform became a foundation for the telecom's ML journey across multiple business units. Once routing optimization proved reliable under live traffic conditions, the same architecture pattern — streaming ingestion, time-series modeling, and automated decision layers — was extended to adjacent operational problems across the organization.

## First Steps

If your organization manages complex multi-party routing, pricing, or logistics decisions across markets, the same pattern applies: high-volume data ingestion, time-series modeling, and production-first delivery.

Start with the highest-cost corridor or the most manual decision process. Build a working system against that single target. Expand once it proves itself in production. If an operational routing workflow is already defined and needs to reach production, [AI Workflow Integration](#) is the direct build path.

## References

1. GSMA. [Roaming](#). *GSMA*, 2024.
2. McKinsey & Company. [How Generative AI Could Revitalize Profitability for Telcos](#). *McKinsey*, 2024.



ML LABS

Custom AI Systems for High-Value Workflows

[mllabs.com](http://mllabs.com)