Every year, hundreds of millions of containers move across oceans and the systems that track them still can't reliably answer the simplest question: when will it actually arrive, and why? We're building the answer from first principles. Over the past 18 months, we have mapped 1,400 ports, 28 global chokepoints, 140 carriers, and 7,500 vessels at a physical level that no commercial dataset comes close to and layered 54 million customs manifests on top of that to create per account, per-lane intelligence that doesn't exist anywhere else. The ML problems we're working on are genuinely hard: we're building explainable prediction models in complex logistics networks, understanding root causes in chaotic real-world systems, and creating the intelligence layer that turns data into decisions for enterprise supply chains. If you've been waiting for a dataset you couldn't have built yourself, a problem space with no solved benchmark to copy, and work that shows up in the real economy the day it ships, this is it.

ML Data Scientist / Deep Learning Scientist

Pref. location: Hyderabad Salary: Negotiable

WHAT THE RIGHT PERSON LOOKS LIKE

This is a production science role, not research-only. We need someone who can carry forward our current iteration of models as new data flows in, as algorithms need re-tuning, and as the model suite expands. We need someone whose models must be running live, being called by real users or systems, and being maintained over time. 4 - 8 years of experience, with both of the following:

Traditional ML:

- Linear regression, logistic regression, tree-based models (XGBoost, LightGBM, CatBoost)
- Understanding structured/tabular models, feature engineering, model evaluation, crossvalidation, and production drift monitoring

Deep Learning:

- Transformer architectures, NLP algorithms - sequence models, attention mechanisms, and time-series deep learning
- Ability to work with and improve existing models - not just use pre-built HuggingFace wrappers

Stack & Tools:

- Python (pandas, scikit-learn, XGBoost, PyTorch or TensorFlow)
- Model registries and experiment tracking (MLflow or equivalent)
- Working knowledge of batch vs real-time inference trade-offs for B2B use cases
- Basic understanding of agentic pipelines - how model outputs are consumed downstream by an agent or API

WHAT YOU'LL OWN

- **Maintain, retrain, and improve our current suite of models around forecasting and explainability**
- **Define the architecture of the agentic pipeline (the ML Engineer builds it; this role defines the logic and what it should produce)**
- **Own the model evaluation framework - tracking accuracy, identifying drift, rerunning models as new data arrives**

WHAT YOU'LL TAKE AWAY FROM THIS EXPERIENCE

You'll be working directly with 18+ months of proprietary model development - physics-based port models, transformer architectures, causal AI - and you'll have the scientific freedom to improve and expand it. The work you do will directly determine whether enterprise supply chains become predictive rather than reactive.