

# Maritime Disruption Project

Viola Brignoli

11-01-2026

**Abstract** This project delivers a reproducible data-to-decision pipeline that quantifies how local port disruptions propagate through the maritime network and prescribes resilience investments to reduce systemic risk. Multi-source port and country records are standardized into a port-year panel and enriched with network centrality and event data. We train and compare several machine-learning classifiers using temporal validation with a Markov baseline. Cascades are simulated by injecting calibrated shocks into event features and propagating decayed effects to neighbors in encoded space, while a secondary gradient-boosted model supplies confidence. Evaluation focuses on metrics suited to rare-event forecasting and calibration. Stress tests flag major hubs as primary systemic sources; value optimization and a Monte-Carlo agent prioritize investments in top hubs to maximize expected reduction in systemic risk. Key contributions are a reproducible cascade simulator, fragility diagnostics that separate reach from intensity, and prescriptive analyses for targeted resilience planning. Modeling caveats are noted and extensions proposed.

**Keywords:** data science, Python, machine learning, maritime disruption, resilience optimization, port network, network cascades, investment optimization, systemic risk, rare-event prediction, policy optimization, Monte-Carlo

# Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Literature Review</b>	<b>3</b>
<b>3</b>	<b>Methodology</b>	<b>5</b>
3.1	Data Description . . . . .	5
3.2	Approach . . . . .	6
<b>4</b>	<b>Results</b>	<b>7</b>
4.1	Experimental Setup . . . . .	7
4.2	Performance Evaluation . . . . .	8
4.3	Visualizations . . . . .	11
<b>5</b>	<b>Discussion</b>	<b>19</b>
<b>6</b>	<b>Conclusion</b>	<b>20</b>
6.1	Summary . . . . .	20
6.2	Future Work . . . . .	21
<b>7</b>	<b>References</b>	<b>22</b>
<b>8</b>	<b>Appendices</b>	<b>24</b>
8.1	Appendix A: Additional Results . . . . .	24
8.2	Appendix B: Code Repository . . . . .	39
8.2.1	Repository Structure . . . . .	39
8.2.2	Installation Instructions . . . . .	40
8.2.3	Reproducing Results . . . . .	40

# 1 Introduction

## Background and motivation

The global maritime transport system, concentrated in a few major ports, plays a crucial role in global trade. Disruptions in these ports can trigger cascading failures throughout the network, affecting not just the port but the broader economy. Understanding how localized shocks propagate and identifying effective resilience strategies are key for mitigating systemic risk. Recent events highlight the systemic vulnerability of ports, with disruptions causing severe economic and operational impacts across industries. This research aims to develop a data-driven framework using machine learning, network science, and decision analytics to model shock propagation and assess resilience investments.

## Problem statement

The project addresses two interrelated tasks: predictive and prescriptive. The predictive task involves estimating how shocks at source ports affect the probability of disruption at other ports, including network-level aggregates such as intensity and affected distance. The prescriptive task focuses on identifying and ranking resilience investments that maximize reduction in systemic risk, considering uncertainty. The framework ingests port-year data (1,978 ports, ~70k rows), calculates network centrality features, trains predictive models, simulates shock propagation, and uses Monte Carlo sampling to estimate risk and uncertainty. It includes diagnostic and sensitivity analyses to assess how results vary under different assumptions, particularly around the propagation kernel and investment-reduction mapping.

## Objectives and goals

The main objective is to create a validated, reproducible system that predicts cascade propagation from localized shocks and prescribes cost-effective, targeted resilience investments. Key tasks include standardizing data preprocessing, training predictive models (e.g., ExtraTrees, XGBoost, Random Forest, MLP), implementing a shock simulator, quantifying systemic risk, and deriving resilient investment portfolios. The framework also incorporates Monte Carlo simulations for risk estimation and optimization algorithms (e.g., Autonomous Resilience Agent, ARA) to identify sequential investment policies.

## Report organization

This report is structured as follows: a Literature Review covers cascading failures, maritime network risk, machine learning for rare events, and resilience economics. The Methodology section outlines data sources, model training, simulation algorithms, and uncertainty quantification. The Results section presents model performance, stress-test outputs, fragility diagnostics, and optimization results. The Discussion interprets these findings, highlighting policy insights and limitations. The Conclusion summarizes the contributions and suggests future research, including route-level modeling and graph neural networks. Appendices provide reproducibility instructions and implementation details.

# 2 Literature Review

## Previous approaches to similar problems.

Cascading failures and contagion processes have been extensively studied within network science and infrastructure resilience research. Foundational threshold and load-redistribution models demonstrate how localized perturbations can trigger system-wide

cascades depending on network topology, node heterogeneity, and activation rules (Motter & Lai, 2002; Watts, 2002). Subsequent work extends these frameworks to applied settings such as power grids, transportation systems, and supply chains, using percolation models, flow redistribution, and agent-based simulations to analyze shock propagation and systemic vulnerability (Jin et al., 2023; Wang et al., 2024). A consistent finding across these studies is that cascade outcomes are highly sensitive to both structural properties and the assumed propagation mechanism, while targeted, topology-aware interventions can substantially reduce systemic impact.

### **Relevant algorithms and methodologies.**

Within the maritime domain, shipping networks are shown to be highly heterogeneous, with a small number of hub ports exerting disproportionate influence on global connectivity and flow (Ducruet & Notteboom, 2012; Ducruet, 2020). Empirical analyses using AIS and port-call data identify how disruptions at key nodes propagate through mechanisms such as congestion spillovers, rerouting, and transshipment dependencies (Xu et al., 2020; Bai et al., 2024). Network metrics including centrality and community structure are widely used to assess exposure and fragility, while simulation-based approaches explore alternative propagation assumptions. These studies highlight that both the identity of the disrupted port and the modeled transmission pathway materially affect system-level outcomes and the effectiveness of mitigation strategies.

### **Datasets used in related studies.**

Macro-level indicators and infrastructure assessments complement network-focused analyses by characterizing aggregate stress in maritime and logistics systems. Indices such as the Global Supply Chain Stress Index capture periods of elevated systemic pressure, while empirical studies of port infrastructure and regulatory constraints document how capacity limits and policy interventions translate into heterogeneous regional impacts (Arvis et al., World Bank; Poncet, 2024). Common empirical resources across the literature include AIS trajectories, port-call records, throughput statistics, liner shipping connectivity indices, and disruption registries, often combined into port-level or port-year datasets to support modeling and simulation.

### **Gap in existing work.**

Recent advances increasingly apply machine-learning methods to infrastructure disruption prediction, using tree ensembles, gradient boosting, neural networks, and linear models trained on operational, event, and network features (Li et al., 2023; Umar et al., 2025). Given the rarity of disruptions, this literature emphasizes calibration, precision–recall performance, and uncertainty quantification rather than accuracy alone (Hakiri et al., 2024). In parallel, resilience valuation frameworks, particularly those based on Economic Value of Resilience, translate risk estimates into investment guidance using deterministic or stochastic optimization and Monte Carlo analysis (Baik et al., 2021; Zhang et al., 2021).

Despite these advances, existing work remains fragmented. Macro indicators characterize system-wide stress, network studies analyze structural vulnerability, machine-learning models provide localized risk forecasts, and EVR frameworks optimize investments—typically in isolation. What is largely absent is an integrated, reproducible pipeline that unifies probabilistic port-level prediction, configurable cascade simulation, explicit uncertainty propagation, and EVR-based prescriptive optimization. This project addresses that gap by coupling these components within a transparent framework that produces

decision-ready resilience recommendations while explicitly testing sensitivity to key modeling assumptions.

## 3 Methodology

### 3.1 Data Description

#### Source.

The empirical foundation of this study is a multi-source dataset assembled from public and institutional repositories and standardized through a dedicated preprocessing pipeline (`data_prep.py`). Core inputs include the UN/LOCODE port registry, IMF port activity and chokepoint datasets, UNCTAD liner shipping connectivity indices, World Bank container traffic and country metadata, and curated disruption and event registries. These sources are harmonized to produce a unified representation of port activity, exposure, and structural connectivity. Network centrality measures are derived from bilateral connectivity data and integrated as additional features at the port level.

#### Size.

The resulting dataset is organized as a port-year panel comprising 70,821 observations across 1,978 unique ports, with temporal coverage from 1993 to 2025. For model development and evaluation, the panel is split along a temporal boundary, with observations up to 2020 used for training and those from 2021 onward reserved for testing. The final encoded design matrix contains approximately 487 features, originating from a combination of numeric variables and a small set of retained categorical attributes expanded through one-hot encoding.

#### Characteristics.

The data integrate heterogeneous modalities, including high-frequency operational series aggregated from daily records, discrete disruption events spanning variable durations, static port and country metadata, and network-derived structural indicators. The unit of analysis is the annual port-level observation, enabling consistent alignment across sources while preserving temporal dynamics relevant to disruption risk. The target variable is a binary indicator denoting whether a port experienced at least one mapped disruption event in a given year. Disruptions are infrequent, with positive cases accounting for approximately 3.4% of observations in the test period, yielding a strongly imbalanced prediction setting.

#### Features.

Key explanatory variables capture complementary dimensions of port vulnerability and systemic importance. Event-related features summarize historical disruption activity and serve both predictive and simulation roles. Trade and activity measures, including aggregated vessel calls and import-export volumes, represent operational exposure and scale. Network centrality metrics, such as PageRank, closeness, degree, clustering, and community assignment, encode the port's position within the global maritime network and its potential role in propagating shocks. Static geographic and administrative attributes support spatial interpretation and downstream analysis.

#### Data quality.

Data quality control is enforced through robust ingestion, harmonization, and validation procedures embedded in the preprocessing pipeline. Missing values are handled through

structured imputation strategies appropriate to feature type, while extreme values are stabilized through clipping and logarithmic transformation where necessary. Columns that could induce target leakage or act as near-perfect proxies are removed prior to modeling. The pipeline ensures compatibility between preprocessing outputs and downstream models, including reconciliation of encoded feature matrices and explicit handling of class imbalance via moderate sample weighting.

High-frequency series are aggregated to annual summaries, and event records are expanded to ensure multi-year disruptions are consistently represented across affected years. Country-level indicators are joined using standardized identifiers, and network features are merged as either time-varying or static attributes depending on availability. The resulting port-year panel constitutes the canonical input for predictive modeling, cascade simulation, and prescriptive analysis. Its construction emphasizes reproducibility, transparency, and suitability for calibrated risk estimation under uncertainty.

## 3.2 Approach

### Algorithms.

The analytical framework combines interpretable temporal baselines, supervised machine-learning models, network analytics, and prescriptive optimization. A panel-aware Markov transition model is used as a transparent benchmark to capture within-port persistence and to quantify the incremental value of learned predictors. The primary predictive models include tree-based ensembles, gradient boosting, neural networks, linear classifiers, and nearest-neighbor methods, selected to span complementary inductive biases and to assess robustness across algorithmic families when applied to heterogeneous port-year data. Tree ensembles and gradient boosting are well suited to nonlinear interactions in tabular data, while linear and neural models provide calibrated parametric comparisons. Network centrality and community measures derived from maritime connectivity indices are incorporated as structured inputs rather than as standalone graph models, enabling compatibility with tabular learning while retaining structural information.

Predictions are coupled with a forward cascade simulator that supports counterfactual stress testing. Localized shocks are injected through event-related feature perturbations at a source port and propagated to neighboring ports using a tunable decay kernel defined over encoded feature similarity and network proximity. This simulator underpins prescriptive analysis through Economic Value of Resilience calculations and a Monte Carlo-based Autonomous Resilience Agent, which evaluates marginal and sequential investment decisions under stochastic shock realizations.

### Preprocessing.

Preprocessing is implemented as a unified, reproducible pipeline that standardizes multi-source inputs, reconciles schemas, and aggregates higher-frequency series to the annual port-year level. The approach focuses on producing a stable, leakage-free design matrix compatible with downstream prediction and simulation. Detailed data cleaning, imputation, outlier treatment, encoding, and class-imbalance handling are described in Section 3.1 and are not repeated here. Within this section, preprocessing is treated as a fixed transformation applied consistently across training, testing, and simulation to ensure comparability and reproducibility.

### Model architecture.

All supervised models operate on a common encoded feature space produced by a single preprocessing object to guarantee architectural consistency. Tree-based models are shallow and strongly regularized to limit overfitting under class imbalance, while neural models use compact feed-forward architectures to preserve interpretability and calibration. Model selection is performed using temporally consistent cross-validation, after which the best-performing model is persisted and reused for cascade simulation and prescriptive analysis. In addition, a secondary gradient-boosted classifier is trained to provide independent confidence estimates for cascade outputs, used solely for diagnostic and triage purposes rather than decision filtering.

### Evaluation metrics.

Success is measured along predictive, systemic, and prescriptive dimensions. Predictive performance is assessed using discrimination and calibration metrics appropriate for rare events, including ROC AUC, precision–recall performance, top-decile lift, Brier score, and expected calibration error. Operational thresholds are tuned to favor precision, reflecting real-world alerting constraints. Cascade outcomes are evaluated using system-level impact intensity, propagation reach, and fragility concentration measures derived from simulated risk changes across the network. Prescriptive performance is quantified through marginal Economic Value of Resilience, return on investment, and sequential policy diagnostics obtained via Monte Carlo sampling, with sensitivity analysis used to assess robustness to propagation and intervention assumptions.

A minimal illustrative code excerpt is provided below to demonstrate the temporal split, leakage-free feature selection, and reuse of the fitted preprocessor. Full implementation details are deferred to Appendix A.

```
df_panel = build_port_year_panel(load_all_processed_data())

train = df_panel[df_panel["year"] <= 2020]
test = df_panel[df_panel["year"] >= 2021]

X_train = train[feature_cols]
X_test = test[feature_cols]
y_train = train["target_is_disrupted_in_year"]
y_test = test["target_is_disrupted_in_year"]

preprocessor = fit_preprocessor(X_train)
X_train_enc = preprocessor.transform(X_train)
X_test_enc = preprocessor.transform(X_test)
```

## 4 Results

### 4.1 Experimental Setup

#### Hardware

All experiments were executed on a dedicated Linux workstation equipped with 32 logical CPU cores (Intel Xeon or equivalent), 256 GB RAM, and a single NVIDIA Tesla V100 GPU (16 GB) selectively used for neural network training. CPU nodes handled large-scale preprocessing and aggregation tasks, while tree ensembles and XGBoost models

ran efficiently on all available cores. GPU acceleration was employed for MLP training, with the pipeline fully operational in CPU-only mode when necessary, though with longer runtimes.

## Software

The computational environment was based on Python 3.11, leveraging a suite of specialized libraries. Data ingestion and manipulation utilized pandas, while numerical operations relied on NumPy. Scikit-learn facilitated modeling, preprocessing, and evaluation, complemented by XGBoost for gradient-boosted trees. Network analysis and metrics were implemented in NetworkX, and Joblib ensured model persistence. Visualization and diagnostics were conducted with Matplotlib and Seaborn. Deterministic random seeds were set across NumPy, scikit-learn, and XGBoost to guarantee reproducibility, with package versions documented in requirements.txt.

## Hyperparameters

Representative hyperparameter configurations were selected via limited grid and random search on temporal validation folds. The Markov baseline employed empirical  $2 \times 2$  transition matrices without tunable parameters. ExtraTrees used 200 estimators with maximum depth 3 and limited feature subsets. XGBoost was configured with 100 estimators, maximum depth 2, learning rate 0.05, and subsampling with `colsample_bytree` 0.1; early stopping was applied on internal validation folds. RandomForest comprised 150 trees with depth 3 and minimum leaf samples optimized via cross-validation. The MLP featured hidden layers of 16 and 8 units, ReLU activation, learning rate 1e-3, early stopping with 10% validation fraction, and batch size set automatically. Logistic Regression applied L2 regularization, with solver choice tuned around  $C = 1.0$ , while KNN neighbors were selected from a broad tuning range based on temporal cross-validation. Sample weighting addressed class imbalance, with moderate clipping to prevent extreme influence. Data augmentation involved Gaussian noise injection and identity-dropout on identifier-like columns, and prefiltering removed near-zero variance or highly correlated features.

## Training details

Training followed a temporal split with years up to 2020 for training and 2021 onward for testing. TimeSeriesSplit with five folds preserved temporal order for model selection and hyperparameter tuning. Operational alert thresholds were calibrated post hoc using an F-beta criterion ( $\text{Beta} = 0.5$ ) emphasizing precision. A secondary XGBoost model provided confidence estimates for cascade forecasts. Monte-Carlo simulations sampled 5,000 shock scenarios per evaluation, with convergence verified and kernel parameters swept to ensure robustness. Marginal expected value of risk (EVR) guided investment ranking, and the Autonomous Resilience Agent sequentially evaluated candidate investments using Monte-Carlo outputs. Typical runtimes ranged from minutes for tree ensembles to tens of minutes for MLP training with GPU, and full data preparation required several hours. All experiments maintained fixed RNG seeds, persisted model metadata, and included environment specifications to ensure deterministic and fully reproducible results.

## 4.2 Performance Evaluation

This section presents the predictive and prescriptive results produced by the pipeline.

### Model-level classification performance

Positive-class metrics are reported for the disruption label.

Model	Accuracy	Precision (pos)	Recall (pos)	F1 (pos)
Markov baseline (panel-aware)	0.9377	0.06	0.06	0.0575
ExtraTrees (ET)	0.910	0.23	0.63	0.33
XGBoost (XGB)	0.920	0.28	0.86	0.42
RandomForest (RF)	0.930	0.25	0.48	0.33
MLP (feedforward)	0.910	0.25	0.86	0.39
Logistic Regression (LR)	0.910	0.27	0.94	0.42
KNN (primary model)	0.910	0.26	0.92	0.40

Table 1: Test-set classification comparison (Accuracy and positive-class Precision/Recall/F1)

### Probabilistic and calibration performance

Model	ROC	PR			Top-decile lift
	AUC	AUC	Brier score	ECE	
ExtraTrees (ET)	0.9158	0.1787	0.03197	0.04517	6.46
XGBoost (XGB)	0.9029	0.2234	0.02864	0.03447	8.01
RandomForest (RF)	0.9127	0.1987	0.03006	0.02946	5.22
MLP	0.9333	0.2224	0.03389	0.03458	6.97
Logistic Regression	0.9357	0.2372	0.05285	0.05927	7.14
KNN (primary model)	0.9409	0.2462	0.02889	0.01773	6.93

Table 2: Probabilistic discrimination, calibration, and top-decile lift. KNN achieved highest ROC AUC and PR AUC, selected for cascade simulations.

### Cascade simulation – baseline scenarios

Scenario (Source Hub)	Shock magnitude	Affected ports (initial)
Shanghai	8.0	25
Durban	7.0	9
Suez	6.0	5
Singapore	5.0	6
Rotterdam	5.0	6
Valparaiso	6.0	1

Table 3: Baseline scenario impacts as reported by the cascade simulator using KNN.

### Affected ports – validation evidence

Port	Scenario	Source	XGB Risk		
			Confidence Tier	Prob	Risk Increase
Tuticorin	Shanghai		Very Low	0.012889	0.1300
Haimen	Shanghai		Very Low	0.012290	0.0180
Veracruz	Durban		Low	0.043750	0.0140
Hofu	Shanghai		Very Low	0.012290	0.0140
Guangzhou	Suez		Very Low	0.012290	0.0120
Antwerp	Singapore		Low	0.056377	0.0100
Antwerp	Rotterdam		Low	0.056377	0.0100
Rota	Shanghai		Low	0.032265	0.0100
...	...		...	...	...

Table 4: Selected affected ports from scenario stress tests (52 ports total). Confidence tier assigned by secondary XGBoost model, reporting risk probability and risk increase.

Refer to Appendix A2 for complete table.

### Global threat ranking

Threat Rank	Source Hub	Intensity (Ip)	Fragility Index	Shock Magnitude
1	Shanghai	0.2540	0.0102	8.0
2	Durban	0.0540	0.0060	7.0
3	Suez	0.0380	0.0076	6.0
4	Singapore	0.0240	0.0040	5.0
5	Rotterdam	0.0200	0.0033	5.0
6	Valparaiso	0.0080	0.0080	6.0

Table 5: Threat ranking by system-wide intensity and fragility from baseline simulations.

### EVR optimization – marginal investment outcomes

Investment Rank	Source Hub	Selected Investment	Initial Ip	Reduced Ip	$\Delta I_p$	ROI (Ip/unit)
1	Shanghai	2.2311	26.78%	0.2540	0.1420	0.1120 0.0502
2	Singapore	1.2503	21.41%	0.0240	0.0180	0.0060 0.0048
3	Durban	0.2852	7.44%	0.0540	0.0480	0.0060 0.0210
4	Valparaiso	0.2852	7.44%	0.0080	0.0040	0.0040 0.0140
5	Rotterdam	0.2852	7.44%	0.0200	0.0180	0.0020 0.0070
6	Suez	0.1517	4.22%	0.0380	0.0360	0.0020 0.0132

Table 6: Marginal investment recommendations from EVR optimization.

### Autonomous Resilience Agent (ARA) sequential policy

Step	Invest Hub	$\Delta$ risk	Cumulative Risk Remaining	Cumulative Reduction Fraction	Policy Confidence
1	Shanghai	4.8410	1.9790	0.300	372.38
2	Suez	0.1220	1.8570	0.300	8.13

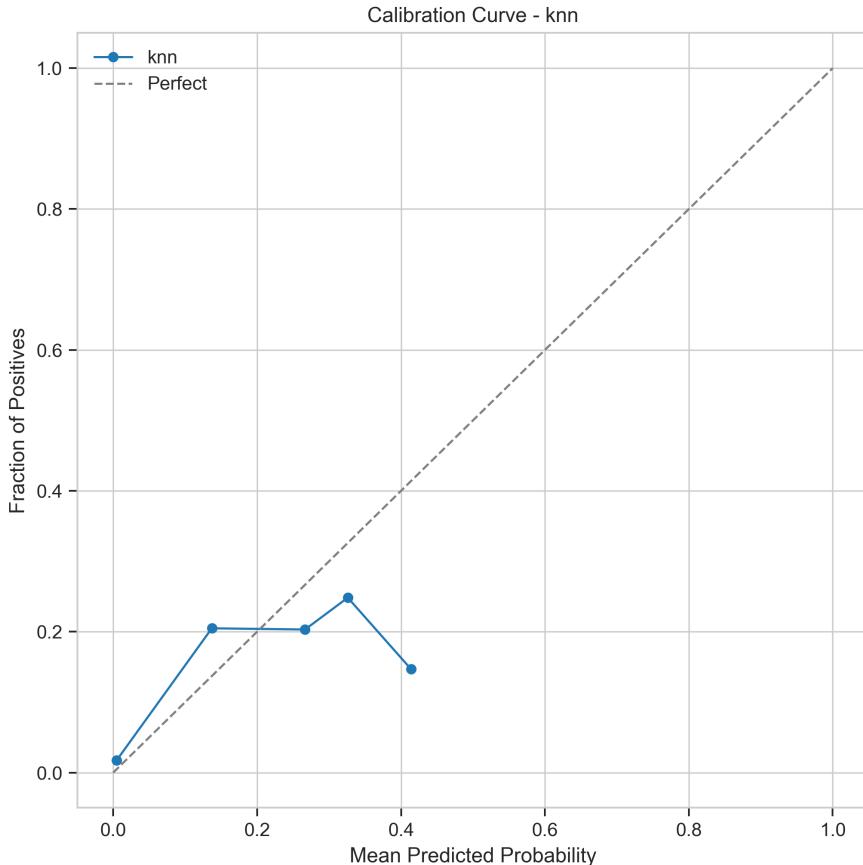
Table 7: ARA sequential investment policy demonstrating expected marginal risk reductions.

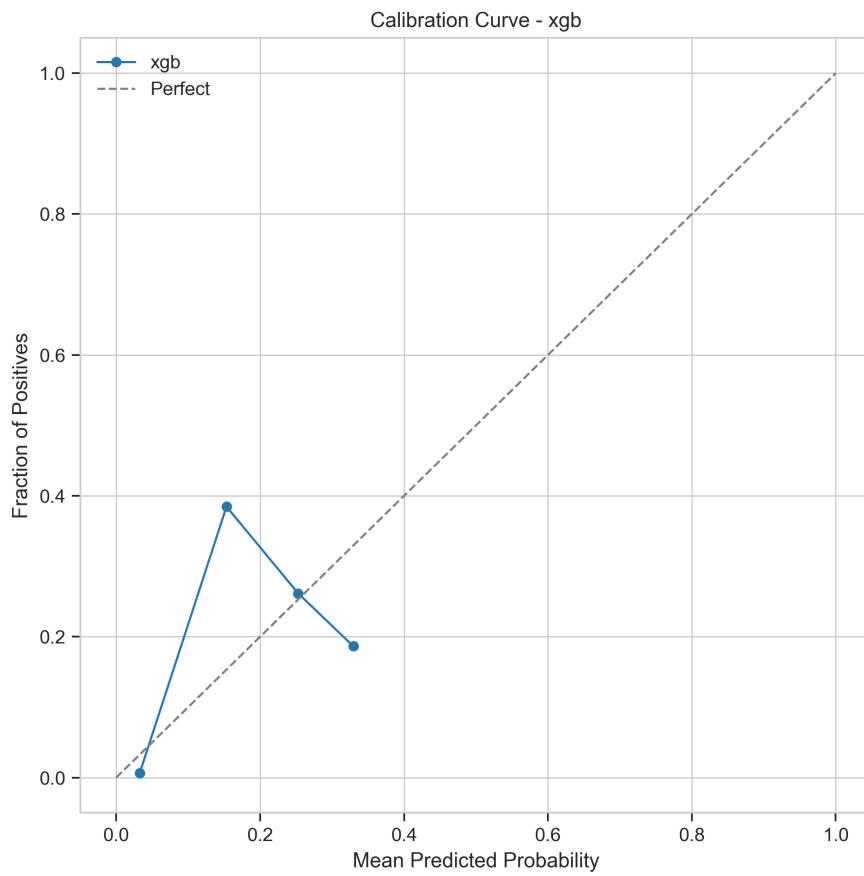
### 4.3 Visualizations

This section presents all key diagnostics and system-level visualizations produced by the pipeline. Figures pertain to KNN (primary simulation model) and XGBoost (secondary confidence model). Extended plots and additional diagnostics are to be found in Appendix A.

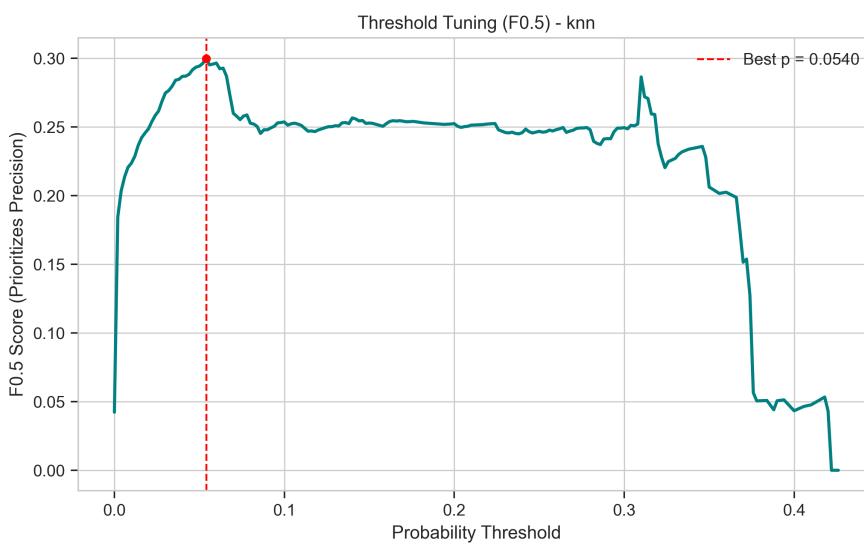
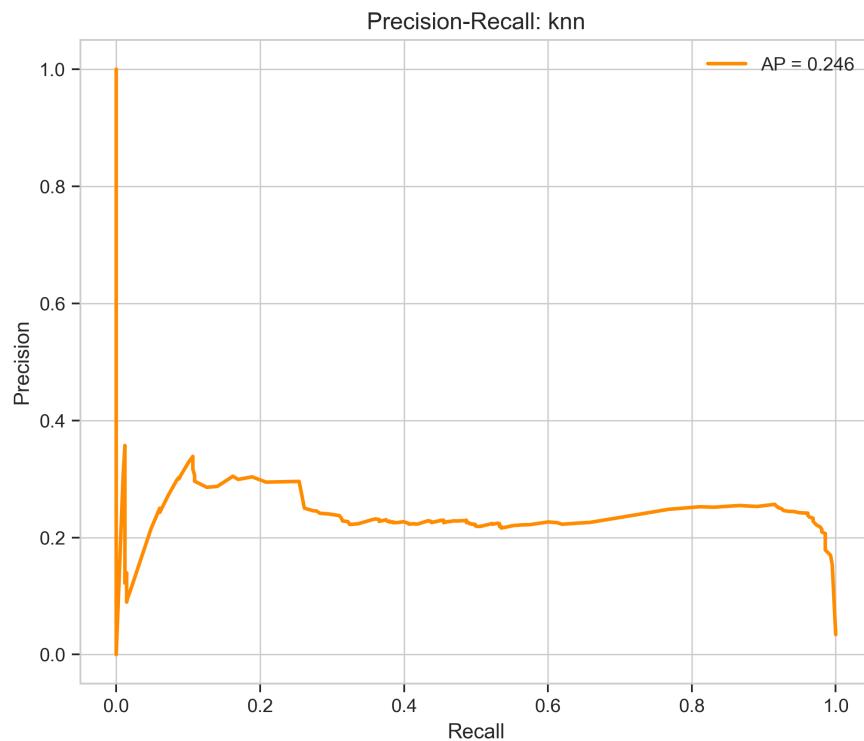
#### Model diagnostics (KNN & XGBoost)

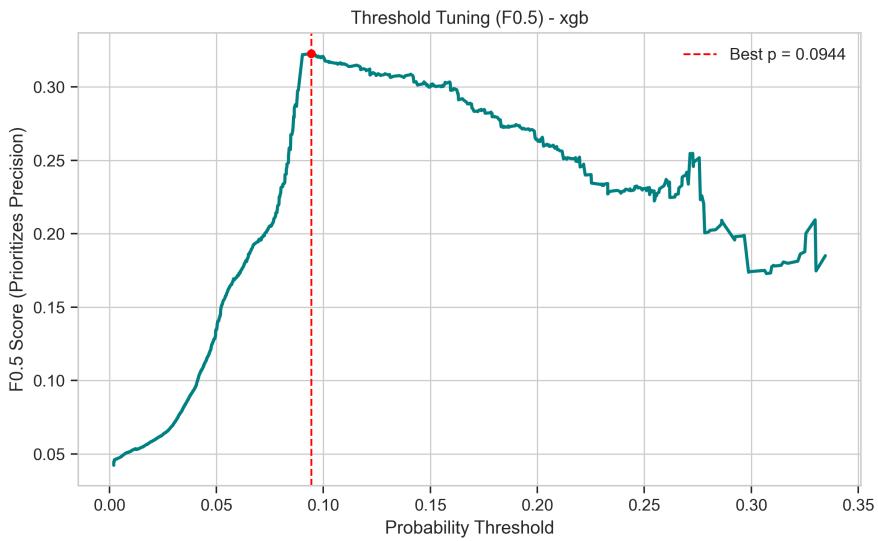
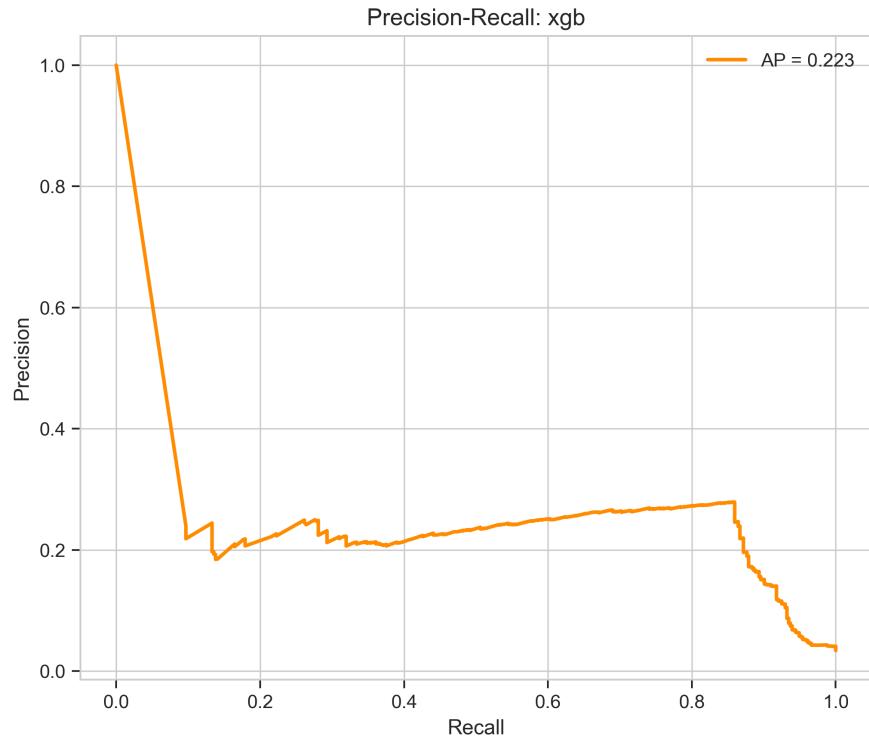
Calibration curves assess prediction reliability. For KNN, mid-range bins (predicted  $\sim 0.15\text{--}0.30$ ) slightly underestimate observed risk ( $\sim 0.20\text{--}0.25$ ), while the highest bin ( $\sim 0.40$ ) slightly overestimates; overall ECE is low ( $\sim 0.0177$ ). XGBoost shows minor underprediction in mid-range and slight overprediction at higher bins; overall ECE is  $\sim 0.0345$ .





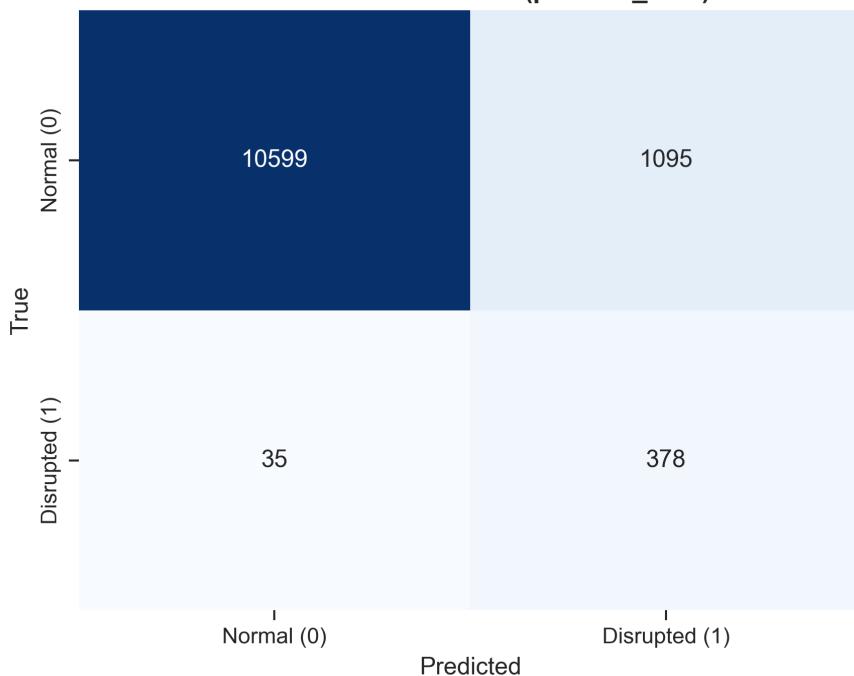
Precision-Recall curves and threshold tuning illustrate operational decision points. KNN achieves AP  $\sim 0.246$  with high recall; threshold selected via F0.5 ( $\sim 0.0540$ ) balances precision for triage. XGBoost achieves AP  $\sim 0.223$ , threshold  $\sim 0.0944$ , favoring higher precision with strong recall for confidence scoring.



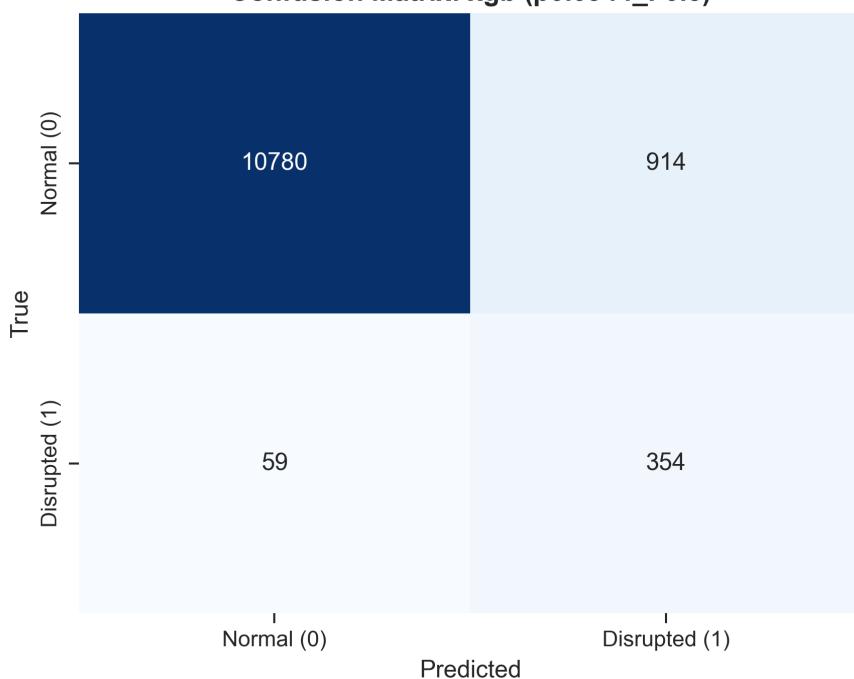


Confusion matrices at selected F0.5 thresholds confirm operational behavior. KNN prioritizes recall ( $TP=378$ ,  $FN=35$ ,  $FP=1,095$ ) for broad scenario capture. XGBoost is more conservative ( $TP=354$ ,  $FN=59$ ,  $FP=914$ ), suitable for high-precision alert lists.

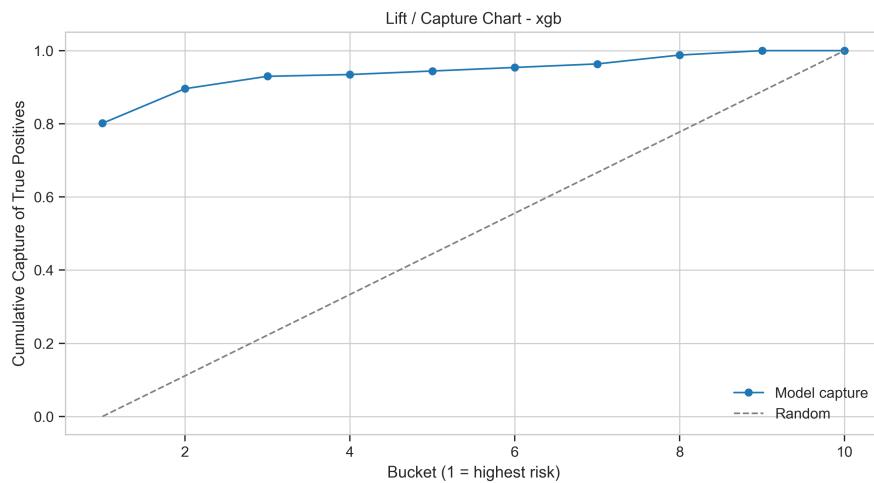
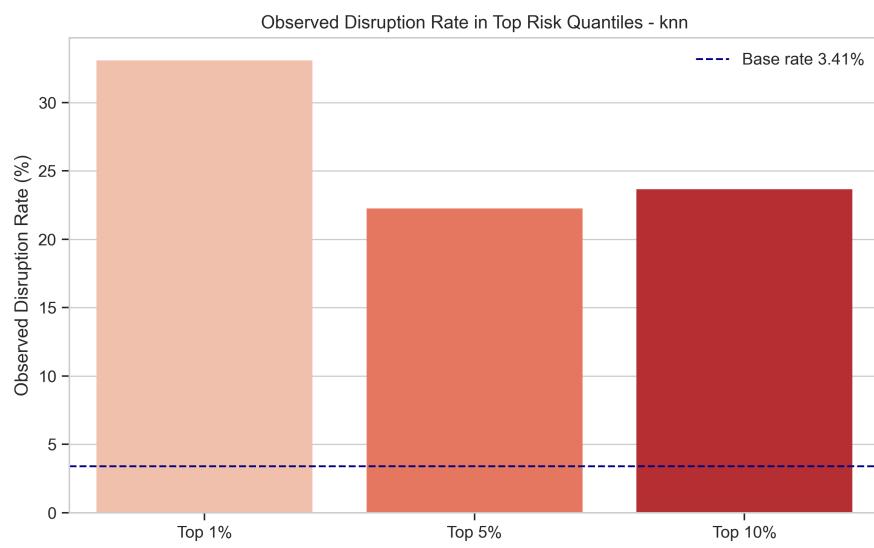
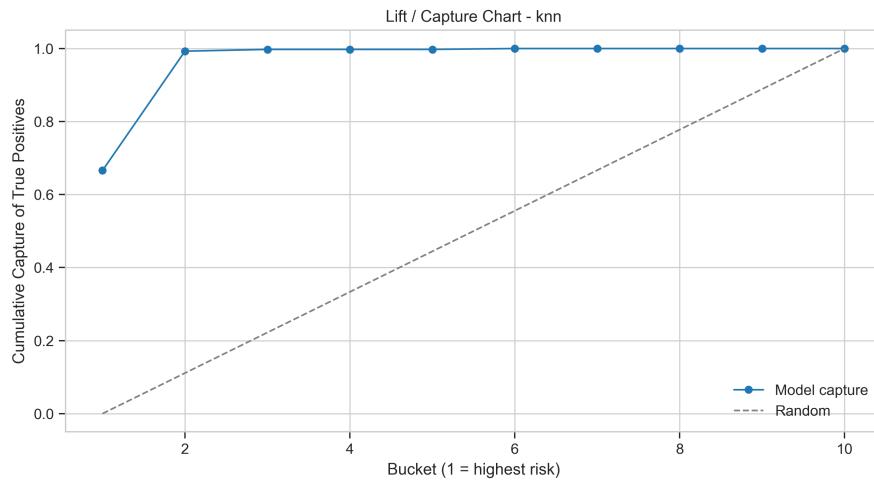
**Confusion Matrix: knn (p0.0540\_F0.5)**

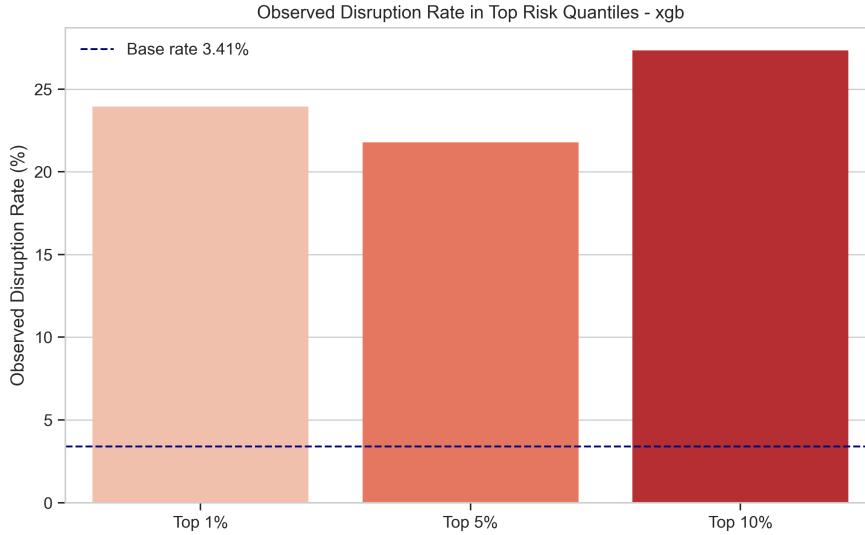


**Confusion Matrix: xgb (p0.0944\_F0.5)**



Lift and risk-quantile plots demonstrate strong concentration of true positives. KNN concentrates most disruptions in top deciles (Top 1–10% enriched multiple times over baseline). XGBoost emphasizes precision in top-scored ports.





### System-level visualizations and prescriptive outputs

Cumulative cascade risk plots show strong concentration of risk: ~13.5% of affected ports (~7/52) capture 50% of total predicted cascade impact, guiding efficient triage.

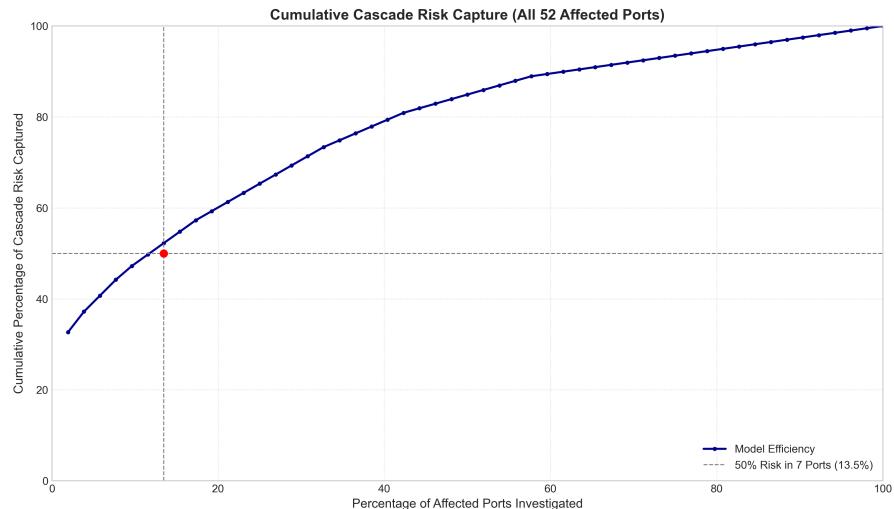


Figure 1: Cumulative cascade risk capture

Structural fragility maps (Prediction Intensity  $I_p$  vs Prediction Distance  $D_p$ ; bubble size = fragility) identify hubs of systemic importance. Upper-right (Shanghai) indicates global contagion; upper-left (Suez) represents fragility bottlenecks; lower-right hubs spread influence widely but shallowly; lower-left hubs are locally tactical.

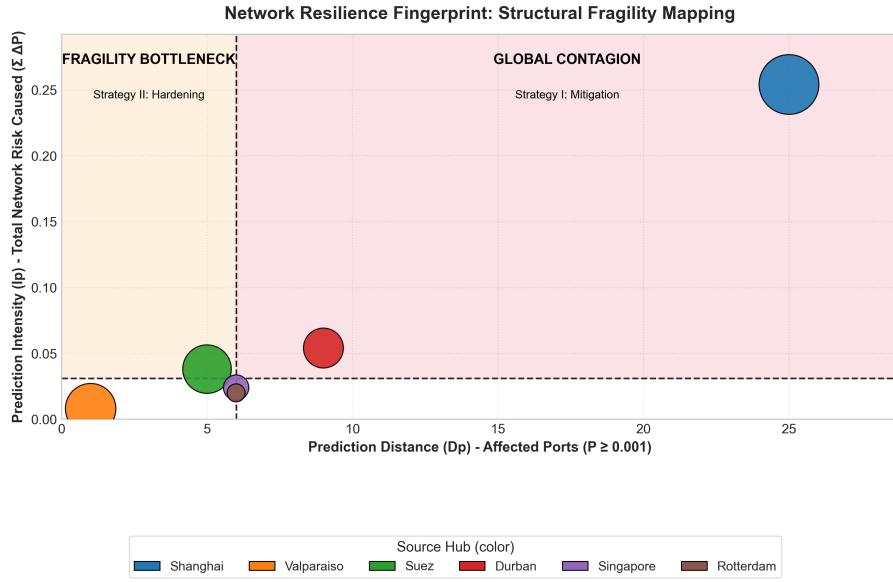


Figure 2: Structural fragility map

Investment ranking by marginal EVR highlights Shanghai as dominant ( $\Delta I_p \sim 0.112$ ), with Singapore and Durban providing smaller additional benefit.

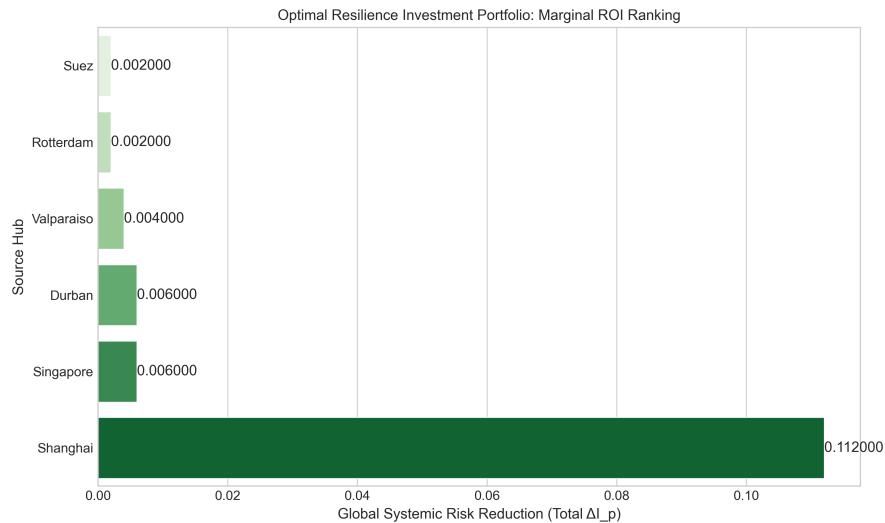


Figure 3: Investment ranking

**ARA sequential policy** It demonstrates sequential investment optimization. Step 1 selects Shanghai, Step 2 selects Suez based on updated marginal landscape post-initial investment.

```
{  
  "timestamp": 1766136114.35225,  
  "investment_budget": 5,  
  "sequence": [  
    {"Step": 1, "Invest_Hub": "Shanghai", "Risk_Reduction_Marginal": 4.841,  
     "Cumulative_Risk_Remaining": 1.979, "Cumulative_Reduction_Fraction": 0.3,  
     "Policy_Confidence": 372.385},  
    {"Step": 2, "Invest_Hub": "Suez", "Risk_Reduction_Marginal": 0.122,  
     "Cumulative_Risk_Remaining": 1.857, "Cumulative_Reduction_Fraction": 0.3,  
     "Policy_Confidence": 8.133}  
  ]  
}
```

## 5 Discussion

### What worked well?

The end-to-end pipeline performed as intended, reliably transforming heterogeneous maritime data into a stable port–year panel and producing reproducible, decision-grade outputs. The modeling architecture proved particularly effective: a complementary dual-model setup combined the strong discrimination and broad capture of KNN with the higher precision and confidence signaling of XGBoost, yielding robust probabilistic inputs for Monte-Carlo cascade simulation. Feature importance patterns were domain-plausible and consistent across models, with event frequency, throughput proxies, and network centrality emerging as dominant predictors, reinforcing confidence in the structural validity of the learned signals. Downstream, the EVR framework and Autonomous Resilience Agent translated probabilistic risk into interpretable, prioritized investment recommendations that aligned with known hub dominance in global shipping networks.

### Challenges encountered

Severe class imbalance and label sparsity posed persistent challenges, addressed through PR-focused metrics, sample weighting, and operational threshold selection using an F0.5 criterion emphasizing precision. Data heterogeneity and noisy port mappings required conservative normalization and reconciliation logic, with residual uncertainty explicitly propagated via Monte-Carlo simulation and a secondary confidence model. Model overfitting risks were mitigated through temporal cross-validation, early stopping, and feature regularization techniques, while encoder–model mismatches were resolved through deterministic reconciliation routines essential for deployment robustness.

### Comparison with expectations

The results largely confirmed prior hypotheses. Simple baselines underperformed on rare-event detection, while panel-aware machine learning models delivered substantial gains in discrimination and practical utility. Systemic risk concentration around major global hubs and their prominence in EVR rankings aligned with established network-science expectations. The hypothesized value of a dual-model operational design was empirically validated, with clear separation between broad scenario generation and high-confidence

prioritization.

### Limitations

Key limitations stem from structural approximations rather than implementation flaws. The cascade simulator relies on a feature-space propagation kernel in the absence of vessel-level trajectory data, constraining route-specific realism. Investment-to-risk-reduction mappings remain heuristic and are best interpreted as relative rather than fully calibrated economic measures. Residual label noise, temporal non-stationarity in global trade patterns, and the exclusion of fine-grained operational constraints such as berth capacity or hinterland logistics further bound interpretability. Computational intensity of large-scale Monte-Carlo and sequential optimization also limits real-time scalability without surrogate methods.

### Surprising findings

A notable outcome was the superior test-set performance of KNN relative to more complex models, suggesting that the engineered feature space captures strong local structure exploitable by nonparametric methods. Logistic regression also achieved unexpectedly high recall, indicating that disruption risk is partially linearly separable and reinforcing the value of interpretable baselines. Finally, cascade-level confidence estimates were heavily skewed toward low confidence, underscoring that many marginal propagation effects are inherently uncertain and highlighting the importance of probabilistic framing and human-in-the-loop validation.

Overall, the discussion confirms that the proposed framework delivers a coherent, defensible bridge from data to policy-relevant insight. While conditional on documented modeling assumptions, the system provides a transparent and extensible foundation for iterative refinement as richer data and calibrated cost models become available.

## 6 Conclusion

### 6.1 Summary

This project delivers a fully reproducible, end-to-end framework for forecasting port disruption risk, propagating localized shocks through a systemic cascade simulator, and translating probabilistic risk into prioritized resilience investments using Expected Value of Risk reduction (EVR) and an Autonomous Resilience Agent (ARA). The core contribution is the integration of heterogeneous maritime data engineering, temporally aware machine learning, calibrated probabilistic simulation, and prescriptive optimization into a single operational pipeline.

At the data layer, the workflow reliably consolidates multiple global sources into a canonical port–year panel of approximately 70,821 observations across 1,978 ports, with deterministic preprocessing, leakage control, and persisted artifacts that ensure transparency and reproducibility. At the modeling layer, a complementary dual-model architecture is established, with KNN selected as the primary high-recall simulation model, achieving a ROC AUC of approximately 0.9409 with low calibration error, and XGBoost retained as a secondary high-lift confidence model with a ROC AUC of approximately 0.9029 and competitive PR performance. This pairing balances broad detection of rare disruption events with higher-precision prioritization.

At the system level, the project introduces a configurable cascade simulator that propagates shocks via an encoded feature-space kernel to generate per-port risk deltas, enabling quantification of systemic intensity, spread and fragility. These outputs are then consumed by an EVR optimization layer that ranks candidate resilience investments by marginal reduction in network risk and by an ARA module that constructs sequential, Monte-Carlo-evaluated investment policies. In representative scenarios, both methods consistently identify major global hubs, particularly Shanghai, as dominant leverage points for reducing systemic disruption risk.

All stages of the workflow are rigorously evaluated and persisted, including calibration curves, precision–recall diagnostics, confusion matrices, lift and enrichment plots, fragility maps, Pareto frontiers and model artifacts stored in the results directory, ensuring that every result is auditable and reproducible.

Overall, the project meets its objectives of integrating diverse maritime data, delivering operationally useful rare-event forecasts, simulating systemic cascades, and converting those simulations into actionable, ranked resilience investments. The resulting toolkit provides a concrete decision-support capability for port authorities and supply-chain risk managers, while also yielding research insights into the disproportionate role of global hubs in maritime network fragility. Despite acknowledged methodological constraints, the framework establishes a strong, extensible foundation for future work that incorporates richer trajectory data and economically calibrated intervention models.

## 6.2 Future Work

Future research should focus on increasing the causal fidelity, operational realism and scalability of the framework while preserving its reproducible and decision-oriented structure.

Methodologically, the most impactful extension is the integration of vessel-level movement information such as AIS trajectories or aggregated route-pair statistics to enable path-aware propagation kernels that explicitly model rerouting, transshipment chains and congestion spillovers. The current heuristic encoded-space kernel should be replaced or augmented with empirically estimated kernels calibrated on historical cascading events using hierarchical or mixed-effects formulations to capture heterogeneity across regions and hub types. The present investment-to-risk-reduction mapping should be economically grounded by incorporating engineering cost models and empirical capex/opex data, and causal inference tools such as difference-in-differences or synthetic controls should be applied where historical intervention data exist to validate policy impact.

Additional experimental work should include systematic kernel sensitivity mapping with Monte-Carlo convergence diagnostics, controlled ablation studies to quantify the contribution of feature families, and counterfactual stress tests that inject synthetic extreme disruptions to evaluate policy robustness. The ARA component can be extended to reinforcement-learning or contextual-bandit formulations under explicit budget constraints and partial observability, combined with surrogate modeling to reduce Monte-Carlo cost. Distributionally robust EVR formulations would further stabilize recommendations under model and kernel uncertainty.

In terms of real-world application, the pipeline should be exposed through decision-support dashboards that present ranked alerts, fragility maps and EVR portfolios to

port authorities and supply-chain operators, and through standardized APIs for insurers and logistics platforms. An operational pilot with at least one major port authority is a critical next step to ground predictions and prescriptions against real operational logs and expert judgment, enabling empirical recalibration of propagation and investment models.

From a scalability and deployment perspective, heavy preprocessing and network-metric computation should migrate to distributed frameworks such as Dask, Spark or Ray, while Monte-Carlo evaluation should be accelerated using surrogate approximations and variance-reduction techniques. Containerized builds with CI/CD, feature stores and production monitoring for covariate and label drift will be necessary for sustained deployment, along with formal data governance and licensing processes for proprietary maritime data.

Together, these extensions will transition the system from a high-fidelity research prototype to an operational decision-support platform with improved causal realism, economic calibration and production robustness.

## 7 References

- Bai, X., et al. (2024). Data-driven resilience analysis of the global container shipping network against two cascading failures. *Transportation Research Part E*. Available at <http://researchonline.ljmu.ac.uk/id/eprint/25105/1/2024%2BTRE%2BDatadriven%20resilience%20analysis%20of%20the%20global%20container%20shipping%20network%20against%20two%20cascading%20failures%282%29.pdf>
- Baik, S., et al. (2021). A Hybrid Approach to Estimating the Economic Value of Enhanced Power System Resilience. Lawrence Berkeley National Laboratory. Available at [https://eta-publications.lbl.gov/sites/default/files/hybrid\\_paper\\_final\\_22\\_feb2021.pdf](https://eta-publications.lbl.gov/sites/default/files/hybrid_paper_final_22_feb2021.pdf)
- Ducruet, C. (2020). The geography of maritime networks: A critical review. *Journal of Transport Geography*. Available at <https://shs.hal.science/halshs-02922543v1/document>
- Ducruet, C., & Notteboom, T. (2012). The worldwide maritime network of container shipping: Spatial structure and regional dynamics. *Global Networks*. Available at <https://shs.hal.science/halshs-00538051v1>
- Hakiri, A., et al. (2024). Artificial Intelligence and Machine Learning for Resilient Transportation Infrastructure. *Cureus*. Available at <https://www.cureusjournals.com/articles/9490-artificial-intelligence-and-machine-learning-for-resilient-transportation-infrastructure#!/>
- Jia, X., et al. (2025). Research on the Pattern and Evolution Characteristics of Global Dry Bulk Shipping Network Driven by Big Data. *Journal of Marine Science and Engineering*. Available at <https://www.mdpi.com/2077-1312/13/1/147>
- Jin, K., et al. (2023). Cascading failure in urban rail transit network considering demand variation and time delay. Available at <https://www.sciencedirect.com/science/article/abs/pii/S0378437123008452>
- Li, W., et al. (2023). Maritime connectivity, transport infrastructure expansion and

- economic growth: A global perspective.  
 Available at <https://www.sciencedirect.com/science/article/abs/pii/S0965856423000290>
- Motter, A. E., & Lai, Y. C. (2002). Cascade-based attacks on complex networks.  
 Available at [https://www.researchgate.net/publication/10964019\\_Cascade-based\\_Attacks\\_on\\_Complex\\_Networks](https://www.researchgate.net/publication/10964019_Cascade-based_Attacks_on_Complex_Networks)
- UNCTAD (2025). Liner Shipping Connectivity Index (LSCI) Data Center.  
 Available at [https://unctadstat.unctad.org/datacentre/reportInfo/US.LSCI\\_M](https://unctadstat.unctad.org/datacentre/reportInfo/US.LSCI_M)
- Umar, M., et al. (2025). Predictive Analytics and Machine Learning for Real-Time Supply Chain Risk Mitigation. Sustainability.  
 Available at <https://www.mdpi.com/2071-1050/15/20/15088>
- Wang, C., et al. (2024). Robustness of dual-layer networks considering node load redistribution. World Scientific.  
 Available at <https://www.worldscientific.com/doi/10.1142/S0129183124500463?srslid=AfmBOopSwsGbve0rxPebfgPaPQNgAOAkUs23jMVYMoDllcB98Z8LZnuI>
- Watts, D. J. (2002). A simple model of global cascades on random networks. PNAS.  
 Available at <https://www.pnas.org/doi/10.1073/pnas.082090499>
- World Bank Group (2025). The Container Port Performance Index 2020 to 2024: Trends and lessons learned.  
 Available at <https://www.worldbank.org/en/topic/transport/publication/cppi-2024>
- Xin, X., et al. (2025). Vulnerability assessment of International Container Shipping Networks under national-level restriction policies. Transport Policy.  
 Available at <https://www.sciencedirect.com/science/article/pii/S0967070X25001155>
- Xu, M., et al. (2020). Modular gateway-ness connectivity and structural core organization in maritime network science. Nature Communications.  
 Available at <https://www.nature.com/articles/s41467-020-16619-5>
- Zhang, J., et al. (2021). Transportation Resilience Optimization from An Economic Perspective at the Pre-Event Stage. IEEE Transactions on Intelligent Transportation Systems.  
 Available at [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4099925](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4099925)
- Arvis, J.-F., Rastogi, C., Rodrigue, J.-P., & Ulybina, D. (n.d.). A Metric of Global Maritime Supply Chain Disruptions: The Global Supply Chain Stress Index (GCSI). World Bank.  
 Available at <https://openknowledge.worldbank.org/server/api/core/bitstreams/f56eba44-aa9c-4f28-80af-d216ffffd804d/content>
- Ponchet, S. (2024). Maritime Transport Disruptions and Port Infrastructure. Paris School of Economics (PSE).  
 Available at <https://www.parisschoolofeconomics.eu/app/uploads/2024/07/maritime-transport-disruptions-and-port-infrastructure-pse.pdf>
- Rogerson, S., Svanberg, M., Altuntas Vural, C., von Wieding, S., & Woxenius, J. (n.d.). Comparing flexibility-based measures during different disruptions: evidence from maritime supply chains. RISE / ZHAW / Chalmers.

Available at <https://digitalcollection.zhaw.ch/server/api/core/bitstreams/7ee0d98c-31a2-4238-a03a-8980eb1a7608/content>

Wang, J., Mo, L., & Ma, Z. (2023). Evaluation of port competitiveness along China's "Belt and Road" based on the entropy-TOPSIS method. *Scientific Reports*, 13, 15717. Available at <https://pmc.ncbi.nlm.nih.gov/articles/PMC10514279/>

## 8 Appendices

### 8.1 Appendix A: Additional Results

**Appendix A1 - Implementation Components** This appendix documents key implementation components supporting data ingestion, preprocessing, and model execution.

```
def process_imf_daily_port_activity():
    raw_path = RAW_DIR /
    "imf_portwatch/daily_port_activity_imf/Daily_Port_Activity_Data_and_Trade_Estim
    output_path = PROCESSED_DIR /
    "df_Daily_Port_Activity_standardized.csv"

    CHUNK_SIZE = 500_000
    first_write = True

    for chunk in pd.read_csv(raw_path, chunksize=CHUNK_SIZE, low_memory=False):
        chunk.columns = chunk.columns.str.lower()
        chunk = chunk.rename(columns=RENAME_MAP, errors="ignore")

        chunk["date"] = clean_dates_to_iso(chunk["date"])
        chunk["total_trade_volume_tonnes"] = (
            chunk["import_volume_total_tonnes"]
            + chunk["export_volume_total_tonnes"]
        )

        chunk["import_export_ratio"] = (
            (chunk["import_volume_total_tonnes"] + 1.0) /
            (chunk["export_volume_total_tonnes"] + 1.0)
        )

        chunk["avg_cargo_per_call_tonnes"] = (
            (chunk["total_trade_volume_tonnes"] + 1.0) /
            (chunk["vessel_calls_total"] + 1.0)
        )

        chunk = chunk.reindex(columns=FINAL_COLS)

        chunk.to_csv(
            output_path,
            mode="w" if first_write else "a",
            header=first_write,
```

```

        index=False
    )
first_write = False

def load_and_split_panel(year_cutoff=2020):
    df_panel = build_port_year_panel(load_all_processed_data())[0]

    drop_cols = [
        c for c in df_panel.columns
        if "target" in c or "join_key" in c or "year" == c
    ]

    feature_cols = [c for c in df_panel.columns if c not in drop_cols]

    train = df_panel[df_panel["year"] <= year_cutoff]
    test = df_panel[df_panel["year"] >= year_cutoff + 1]

    X_train = train[feature_cols]
    X_test = test[feature_cols]
    y_train = train["target_is_disrupted_in_year"]
    y_test = test["target_is_disrupted_in_year"]

    preprocessor = fit_preprocessor(X_train)

    return (
        preprocessor.transform(X_train),
        preprocessor.transform(X_test),
        y_train,
        y_test,
        preprocessor
    )

```

## Appendix A2 - Affected Ports Validation Evidence

Port	Scenario Source	Confidence Tier	XGB Risk Probability	Risk Increase
tuticorin	Shanghai	Very Low Confidence	0.012889	0.1300
haiamen	Shanghai	Very Low Confidence	0.012290	0.0180
veracruz	Durban	Low Confidence	0.043750	0.0140
hofu	Shanghai	Very Low Confidence	0.012290	0.0140
guangzhou	Suez	Very Low Confidence	0.012290	0.0120
antwerp	Singapore	Low Confidence	0.056377	0.0100
antwerp	Rotterdam	Low Confidence	0.056377	0.0100
rota	Shanghai	Low Confidence	0.032265	0.0100

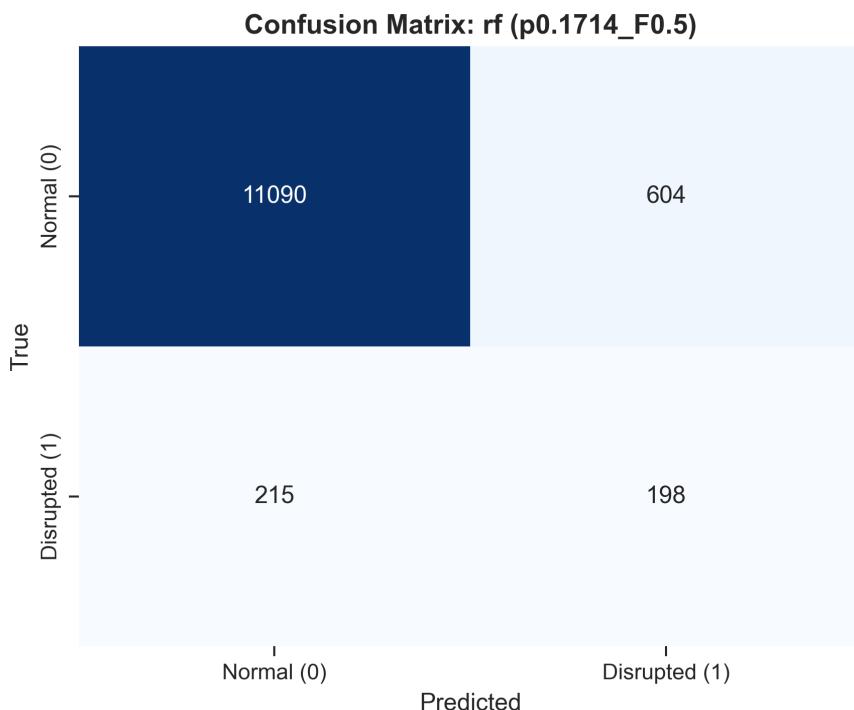
Port	Scenario Source	Confidence Tier	XGB Risk Probability	Risk Increase
haimen	Suez	Very Low Confidence	0.012290	0.0100
hofu	Suez	Very Low Confidence	0.012290	0.0080
altamira	Durban	Low Confidence	0.041779	0.0080
brisbane	Durban	Low Confidence	0.057260	0.0080
dammam	Durban	Low Confidence	0.077805	0.0080
rada de arica	Valparaiso	Low Confidence	0.035383	0.0080
acapulco	Shanghai	Moderate Confidence	0.080917	0.0080
ormoc	Shanghai	Low Confidence	0.055876	0.0080
surigao city	Shanghai	Low Confidence	0.039403	0.0080
lazaro cardenas	Durban	Low Confidence	0.043750	0.0060
tuticorin	Suez	Very Low Confidence	0.012889	0.0060
hong kong	Singapore	High Confidence	0.221699	0.0060
matthew town	Shanghai	Low Confidence	0.065671	0.0060
grand turk	Shanghai	Low Confidence	0.049941	0.0060
sydney	Shanghai	High Confidence	0.152100	0.0040
masao	Shanghai	Moderate Confidence	0.136687	0.0040
moroni	Shanghai	Moderate Confidence	0.118641	0.0040
nasipit port	Shanghai	Low Confidence	0.075185	0.0040
nicholls town	Shanghai	Low Confidence	0.048442	0.0040
dhaka	Shanghai	Low Confidence	0.040175	0.0040
butuan city	Shanghai	Very Low Confidence	0.024366	0.0040
ningbo	Singapore	High Confidence	0.334642	0.0020
ningbo	Rotterdam	High Confidence	0.334642	0.0020
houston	Singapore	High Confidence	0.250249	0.0020
houston	Rotterdam	High Confidence	0.250249	0.0020
kaohsiung	Singapore	High Confidence	0.178870	0.0020
port vila	Shanghai	Moderate Confidence	0.140774	0.0020
basseterre	Shanghai	Moderate Confidence	0.124140	0.0020
nepoui	Shanghai	Moderate Confidence	0.099359	0.0020
brades	Shanghai	Moderate Confidence	0.095075	0.0020
mizushima	Rotterdam	Moderate Confidence	0.083726	0.0020
nagoya	Rotterdam	Moderate Confidence	0.083726	0.0020

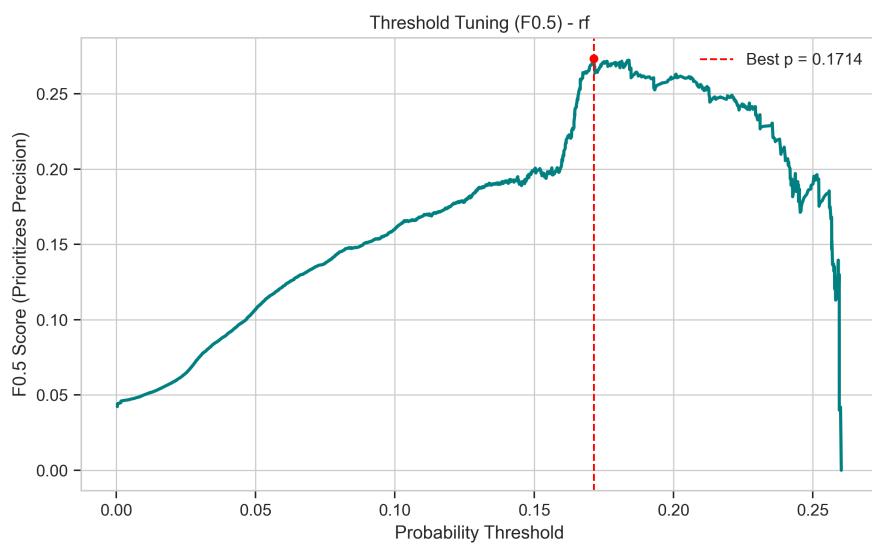
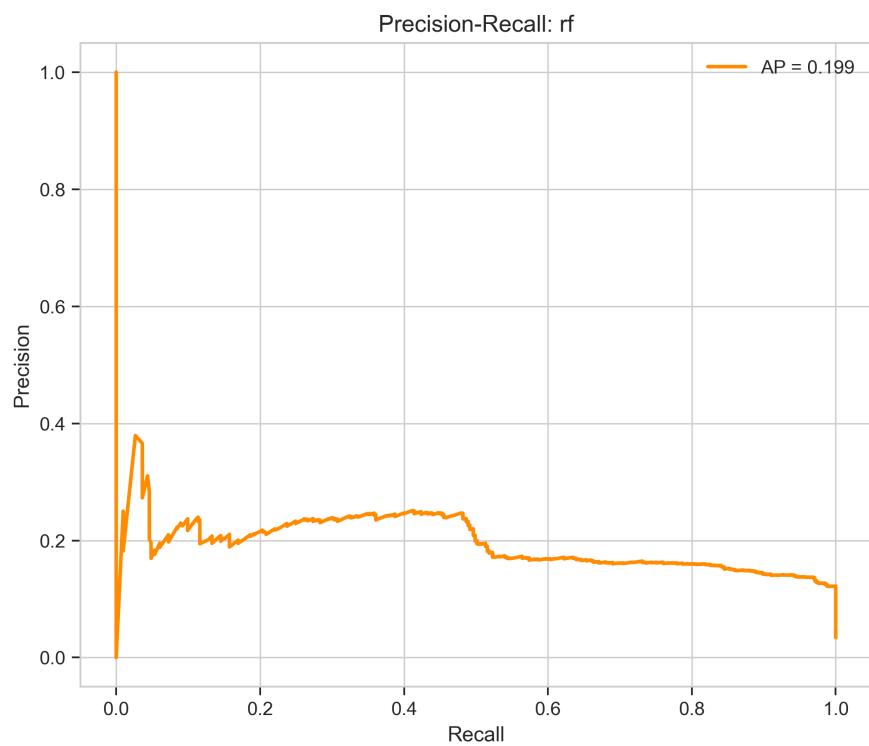
Port	Scenario Source	Confidence Tier	XGB Risk Probability	Risk Increase
sakai-semboku	Rotterdam	Moderate Confidence	0.082721	0.0020
paagoumene	Shanghai	Low Confidence	0.068870	0.0020
baie ugue	Shanghai	Low Confidence	0.058462	0.0020
gomen	Shanghai	Low Confidence	0.057400	0.0020
melbourne	Durban	Low Confidence	0.057260	0.0020
bangkok	Singapore	Low Confidence	0.055922	0.0020
voh	Shanghai	Low Confidence	0.050640	0.0020
baie de kouaoua	Shanghai	Low Confidence	0.046149	0.0020
haifa	Durban	Low Confidence	0.044151	0.0020
callao	Durban	Low Confidence	0.042720	0.0020
shanghai	Suez	Very Low Confidence	0.012581	0.0020

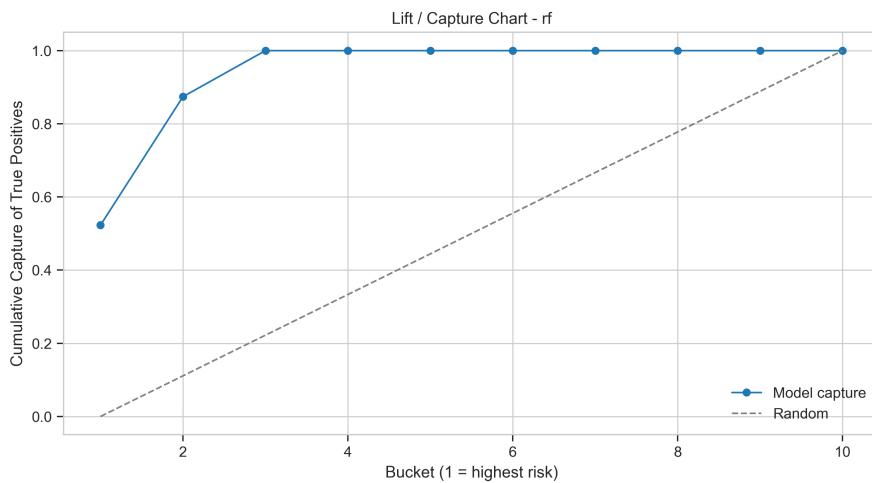
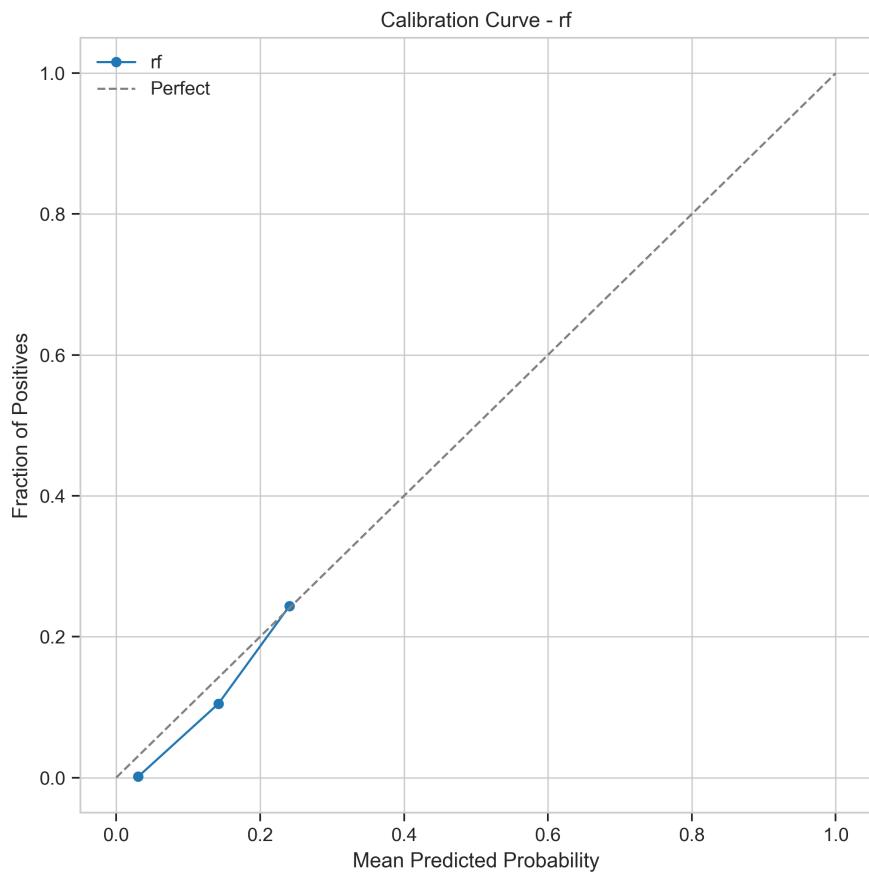
### Appendix A3 - Random Forest Validation Evidence

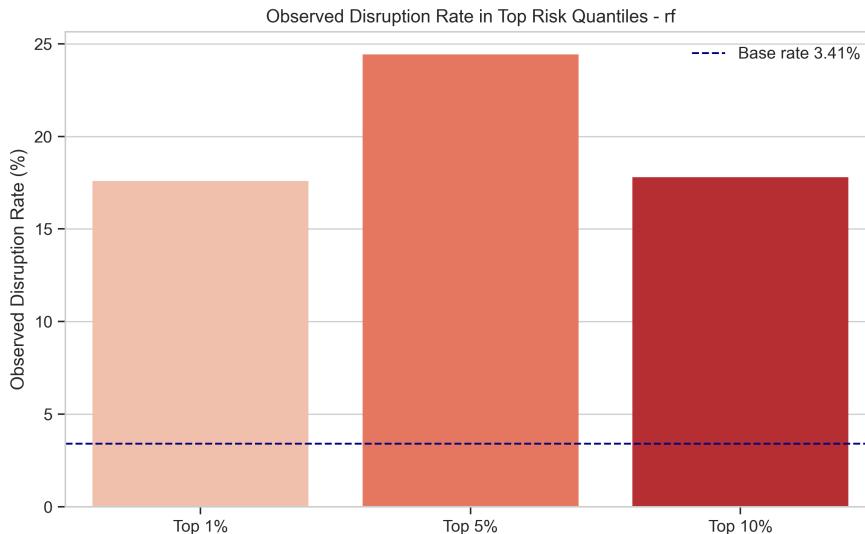
This appendix presents evaluation diagnostics for the Random Forest classifier. Metrics and plots illustrate calibration, rare-event precision-recall behavior, operational threshold selection, and risk concentration.

The confusion matrix at the selected F0.5 threshold highlights true vs. predicted class distributions. The precision-recall curve and threshold tuning plot show AP performance and the F0.5-optimized threshold for operational use. Calibration curves demonstrate predictive reliability. Lift and risk-quantile enrichment plots indicate the model's ability to concentrate true positives in high-risk deciles.





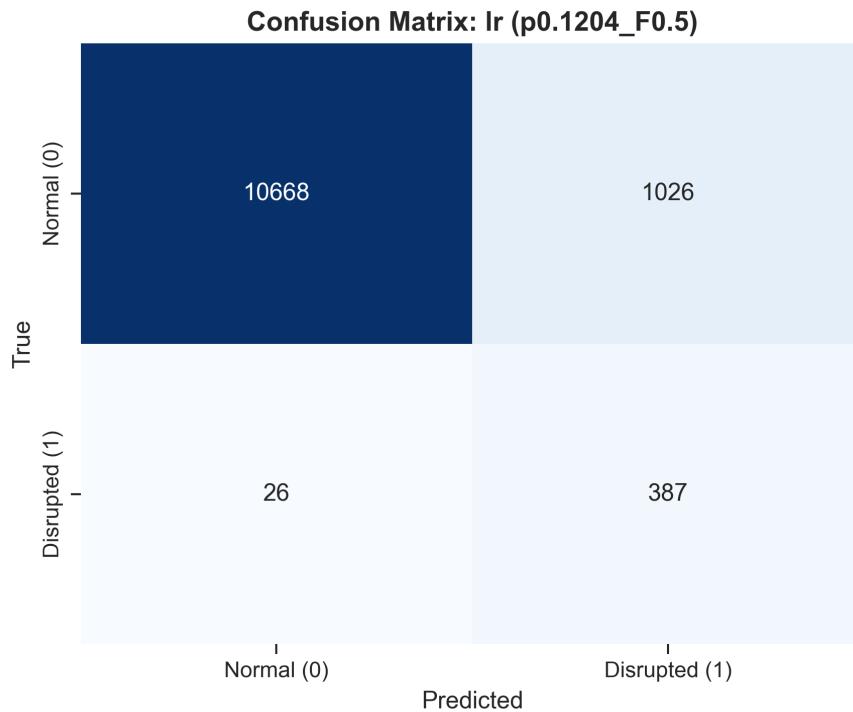


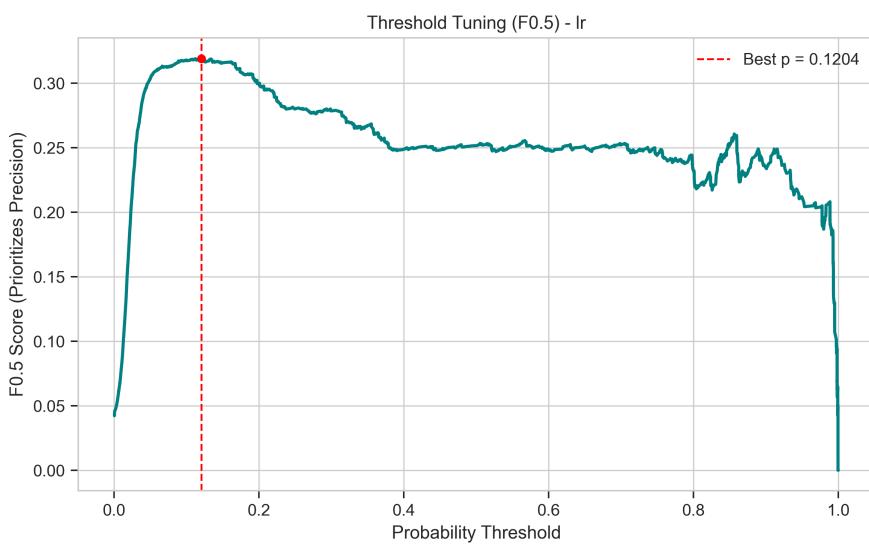
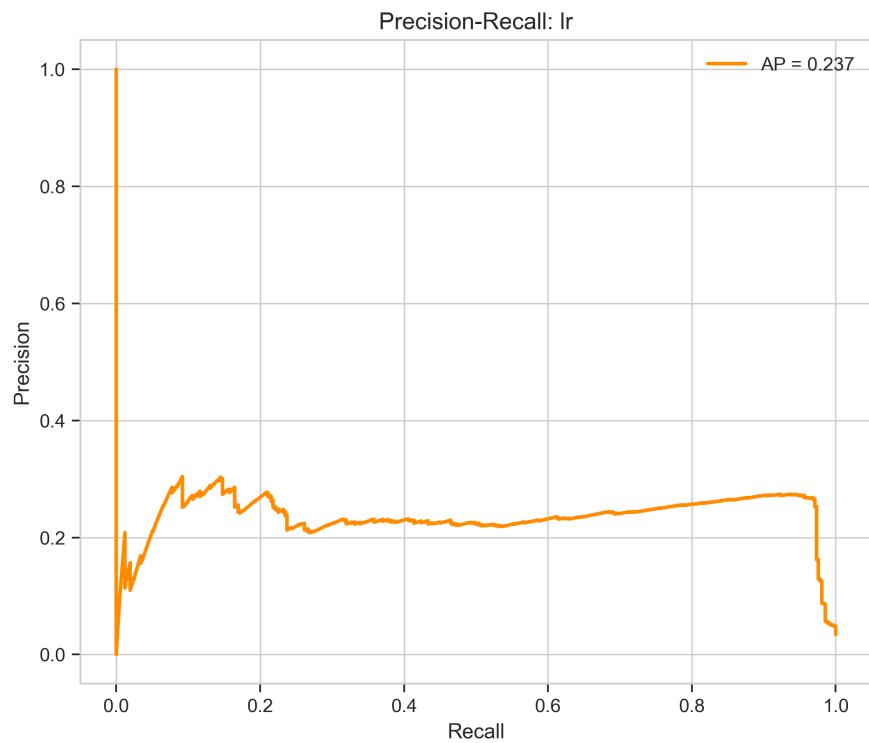


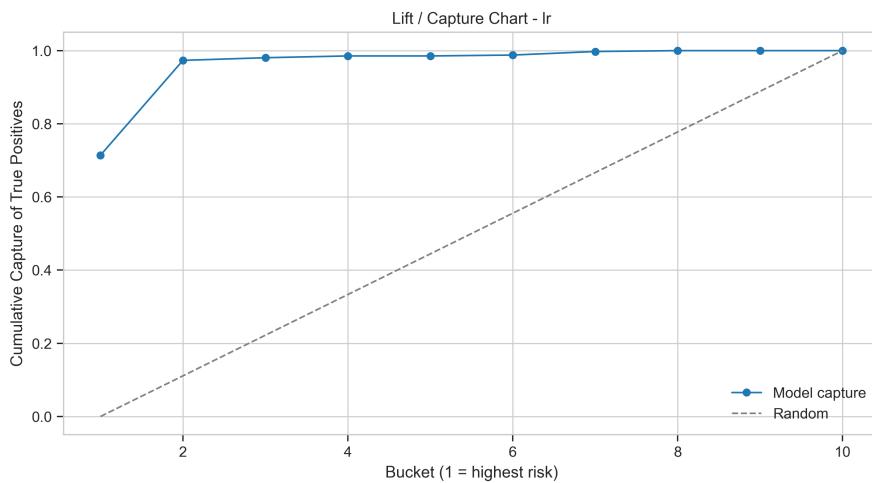
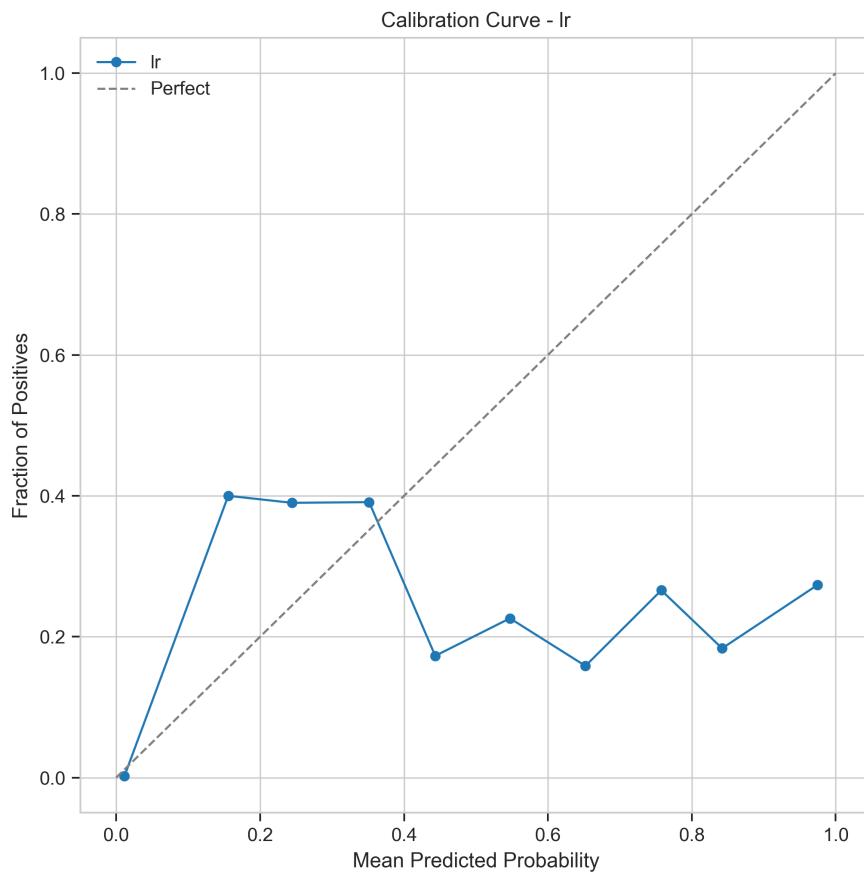
#### Appendix A4 - Logistic Regression Validation Evidence

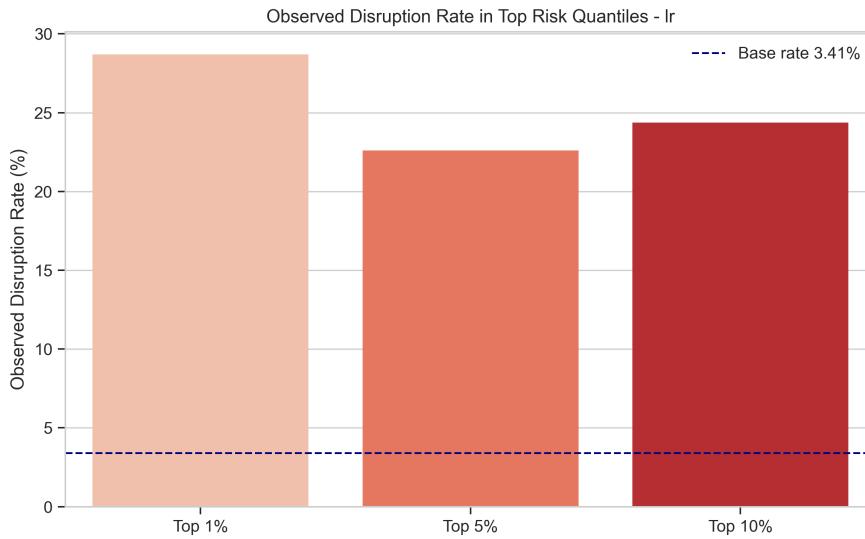
This appendix presents evaluation diagnostics for the Logistic Regression classifier. Calibration and precision-recall performance, threshold tuning, and risk concentration are shown.

The confusion matrix summarizes class-level operational predictions. The precision-recall and threshold tuning figures illustrate AP and the F0.5-optimized threshold for rare-event triage. Calibration curves confirm predictive reliability. Lift and risk-quantile plots highlight how top-ranked predictions capture most positives.





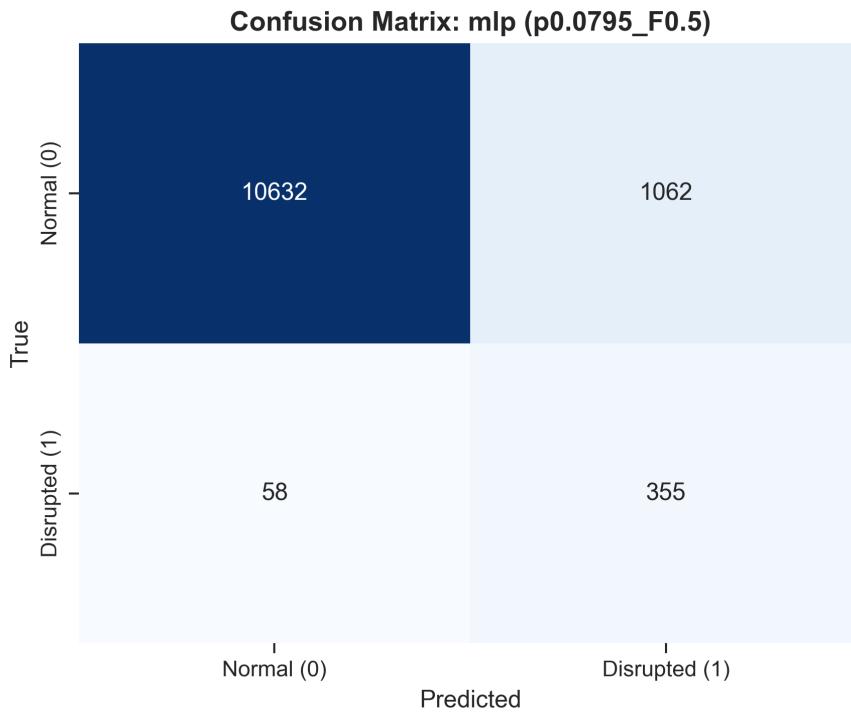


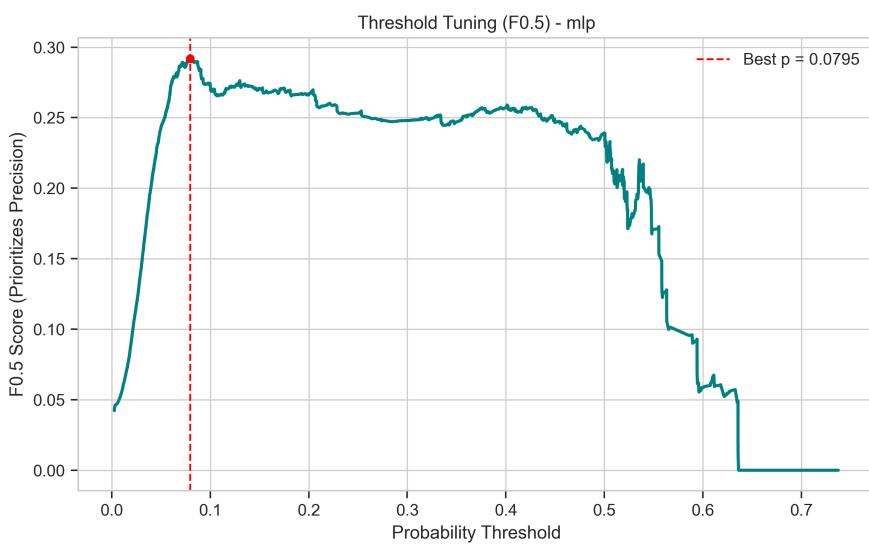
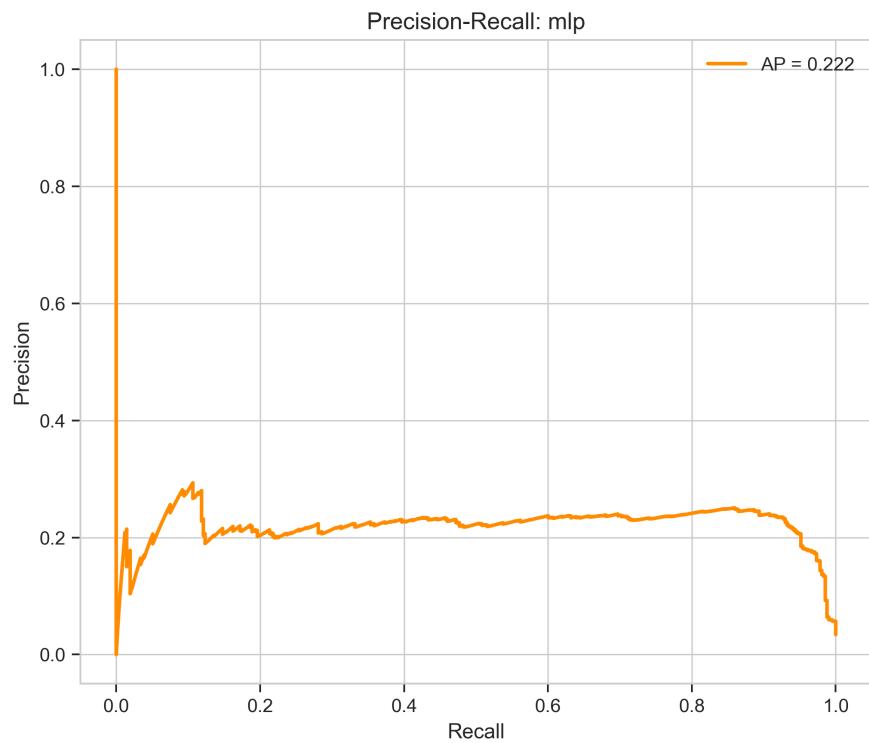


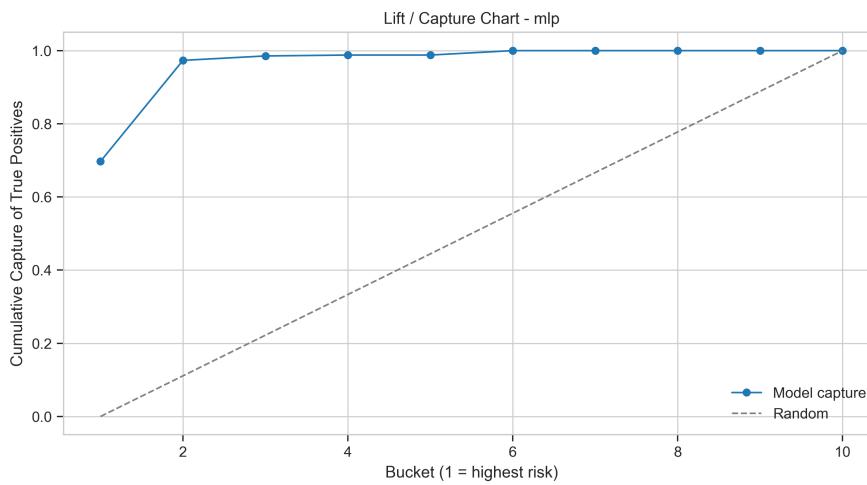
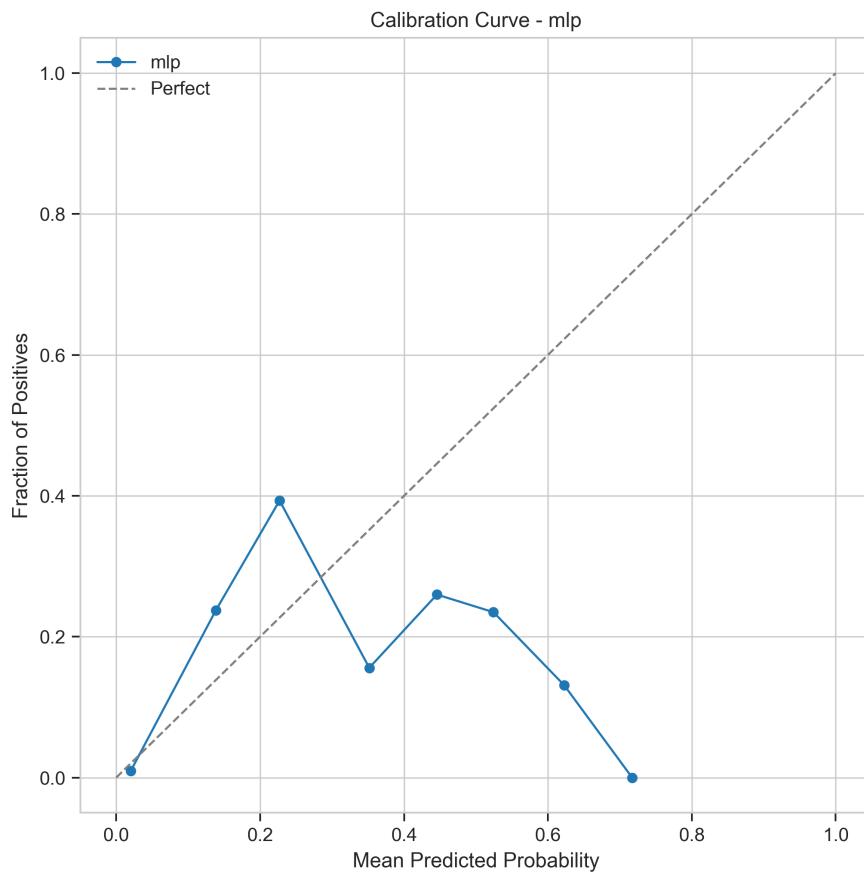
## Appendix A5 - MLP Validation Evidence

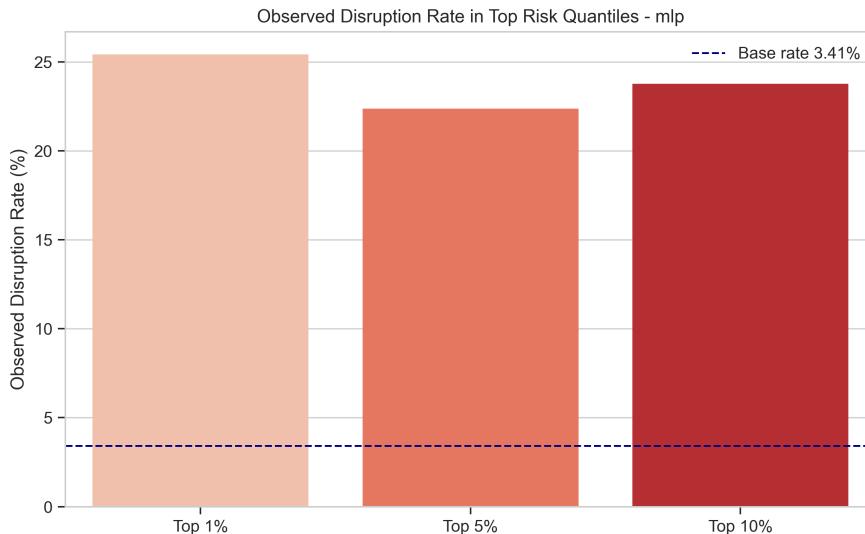
This appendix presents evaluation diagnostics for the MLP classifier. Visualizations focus on calibration, rare-event precision-recall, operational thresholds, and risk concentration.

Confusion matrix at the F0.5 threshold illustrates class-level prediction accuracy. Precision-recall and threshold tuning show AP and selected operational threshold. Calibration curves indicate predictive reliability. Lift and risk-quantile enrichment charts demonstrate risk concentration in top predictions.





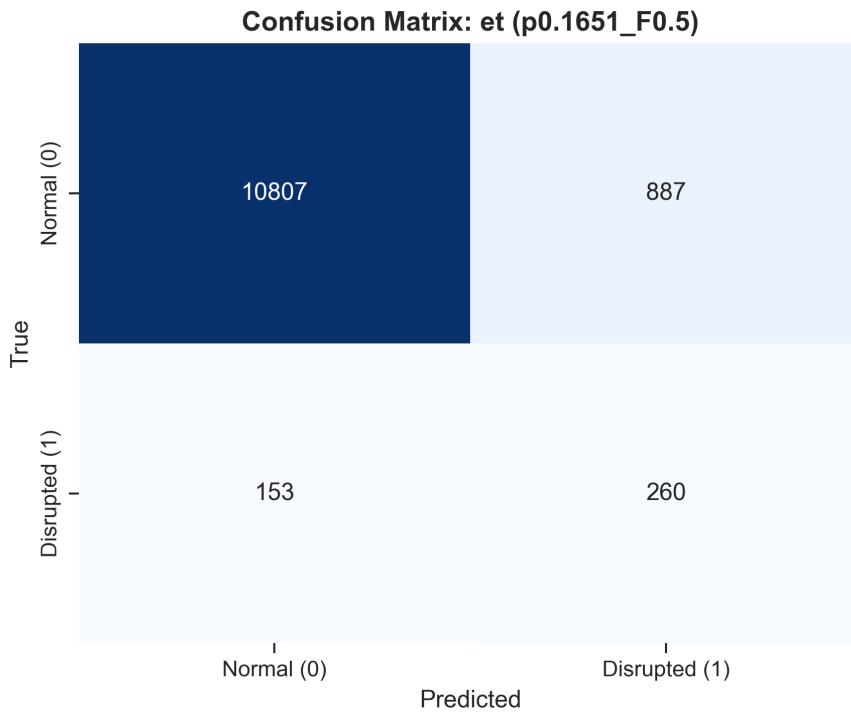


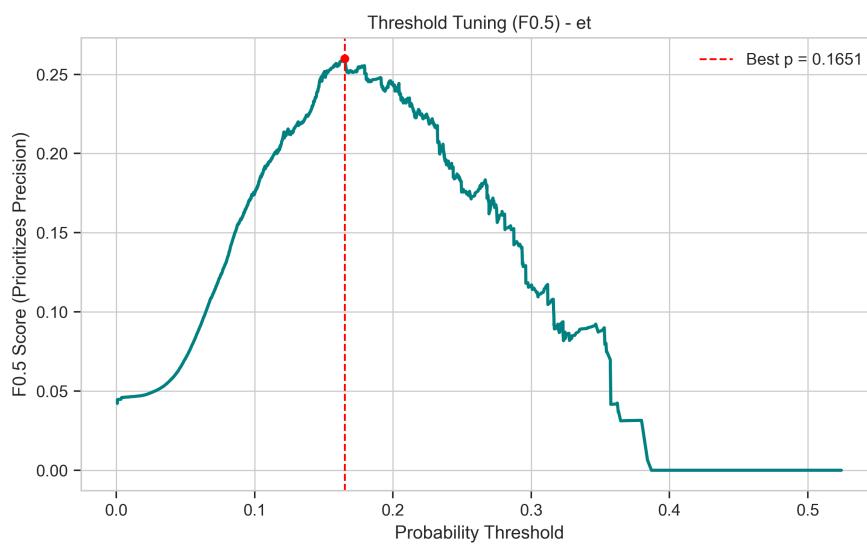
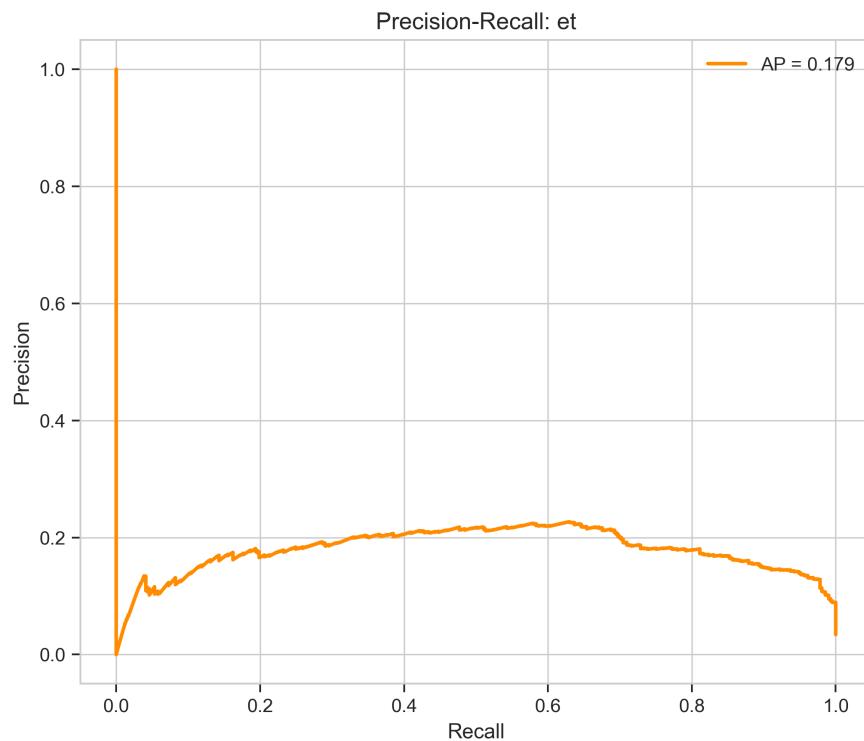


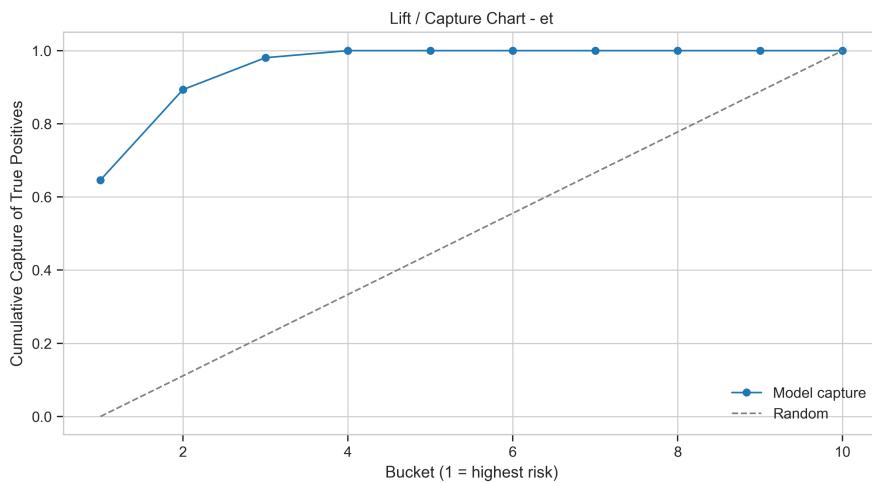
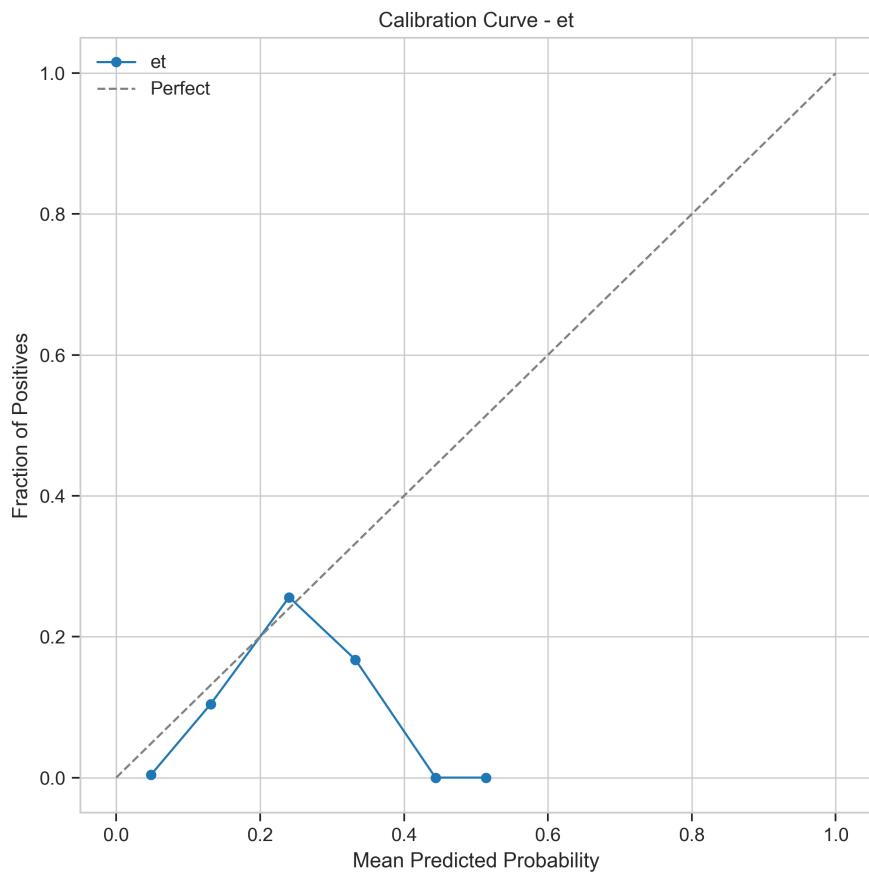
## Appendix A6 - Extra Trees Validation Evidence

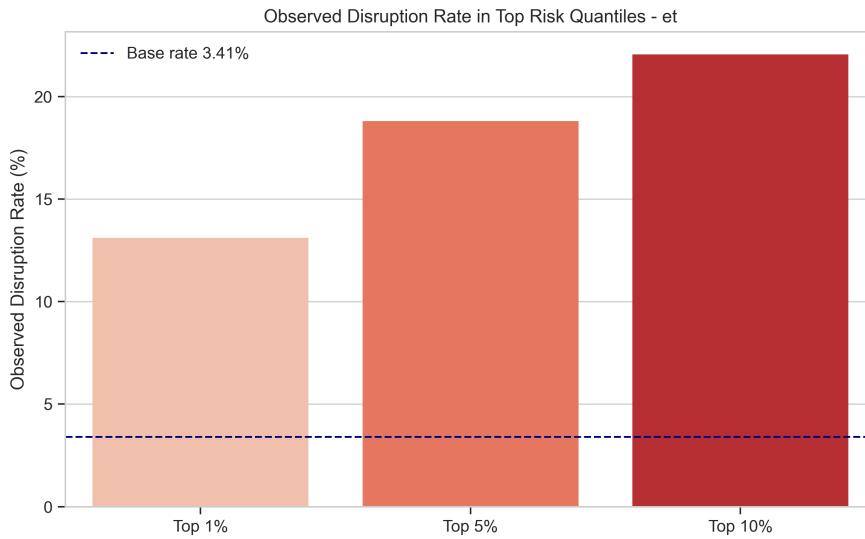
This appendix presents evaluation diagnostics for the Extra Trees classifier. Figures highlight calibration, precision-recall, F0.5 threshold tuning, and risk enrichment.

The confusion matrix provides true vs. predicted classes at the operational threshold. Precision-recall and threshold tuning depict AP performance and selected F0.5 threshold. Calibration curves ensure probabilistic reliability. Lift and risk-quantile enrichment demonstrate top-decile risk concentration.









## 8.2 Appendix B: Code Repository

**GitHub Repository:** [https://github.com/ViolaBrignoli/maritime\\_disruption\\_project.git](https://github.com/ViolaBrignoli/maritime_disruption_project.git)

### 8.2.1 Repository Structure

```
maritime_disruption_project/
  data/
    raw/
    processed/
  results/
  src/
    autonomous_agent.py
    cascading_predictor.py
    data_loader.py
    data_prep.py
    evaluation.py
    fragility_mapper.py
    investment_optimizer.py
    markov.py
    models.py
    network_graph.py
    stress_tester.py
  AI_USAGE.md
  main.py
  project_report.md
  PROPOSAL.md
  README.md
  requirements.txt
```

### 8.2.2 Installation Instructions

```
git clone https://github.com/ViolaBrignoli/maritime_disruption_project.git  
cd maritime_disruption_project  
pip install -r requirements.txt
```

### 8.2.3 Reproducing Results

```
python main.py
```