

Model Evaluation Report

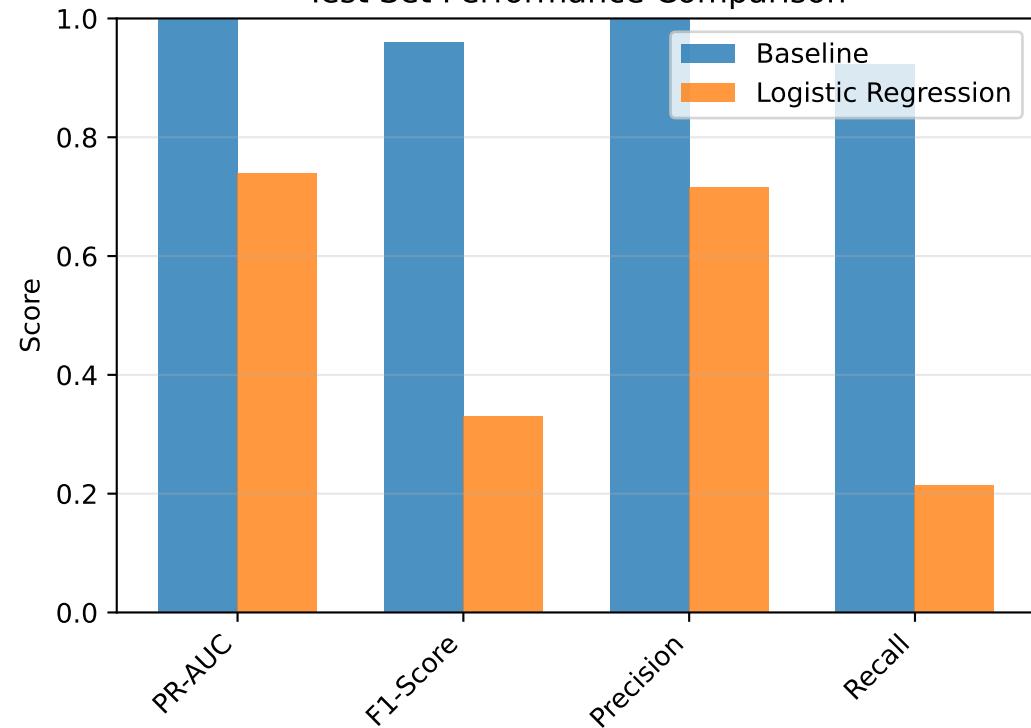
Crypto Volatility Spike Detection

Generated: 2025-11-14 12:35:55

Milestone 3: Modeling, Tracking, Evaluation

Model Performance Comparison

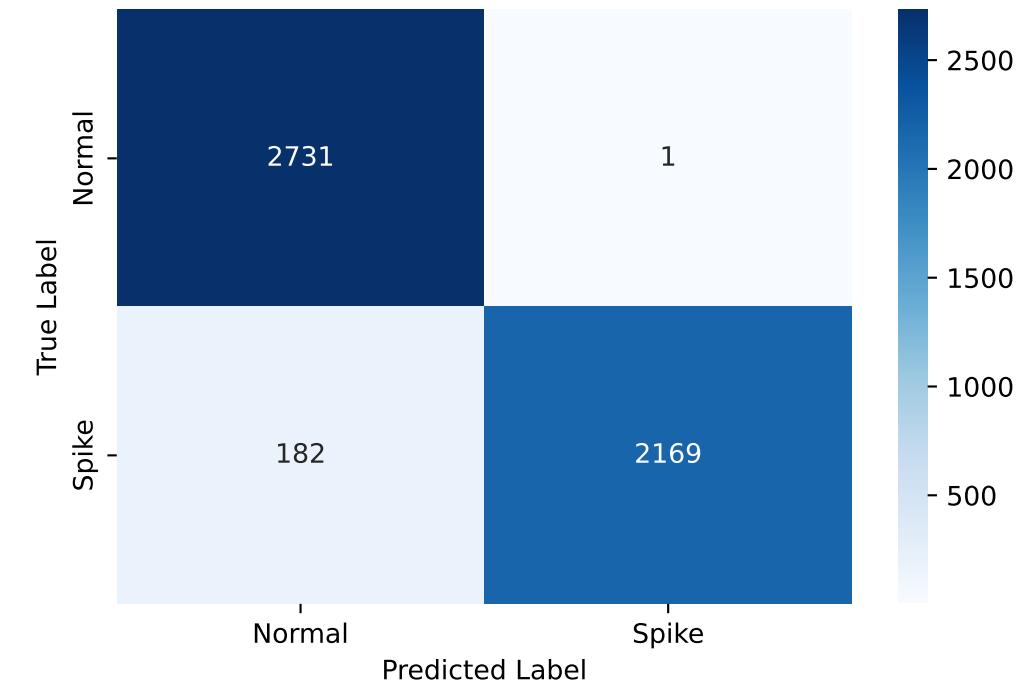
Test Set Performance Comparison



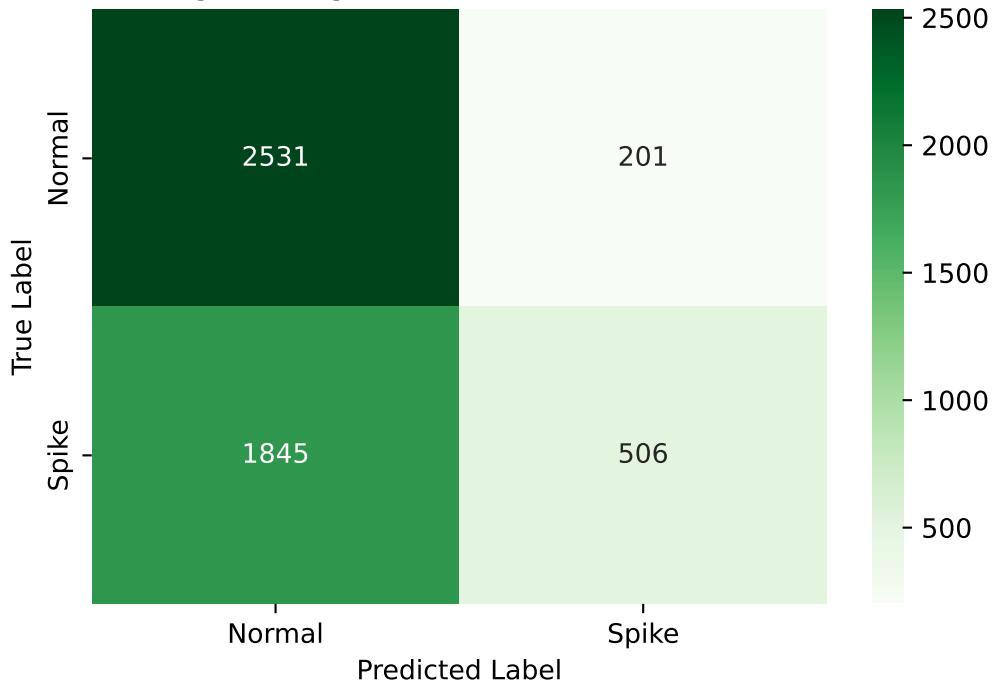
Test Metrics Comparison Table

Metric	Baseline	Log. Reg.
PR-AUC	0.9997	0.7398
F1-Score	0.9595	0.3309
Precision	0.9995	0.7157
Recall	0.9226	0.2152

Baseline - Confusion Matrix



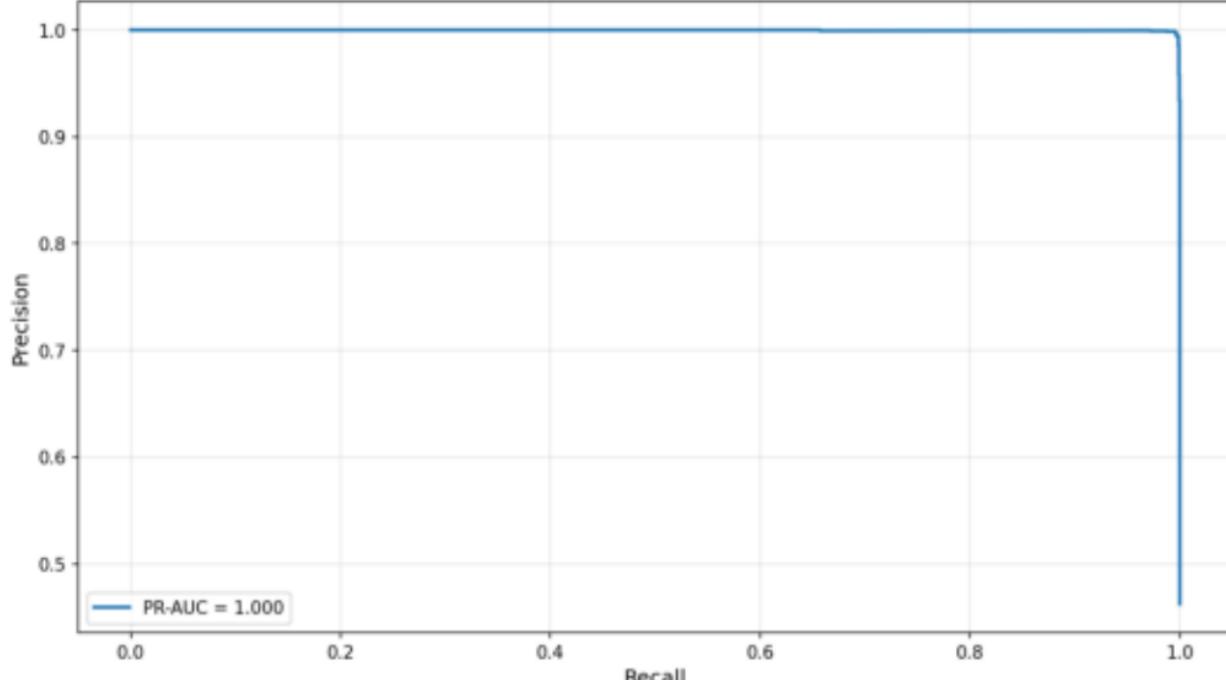
Logistic Regression - Confusion Matrix



Precision-Recall Curves

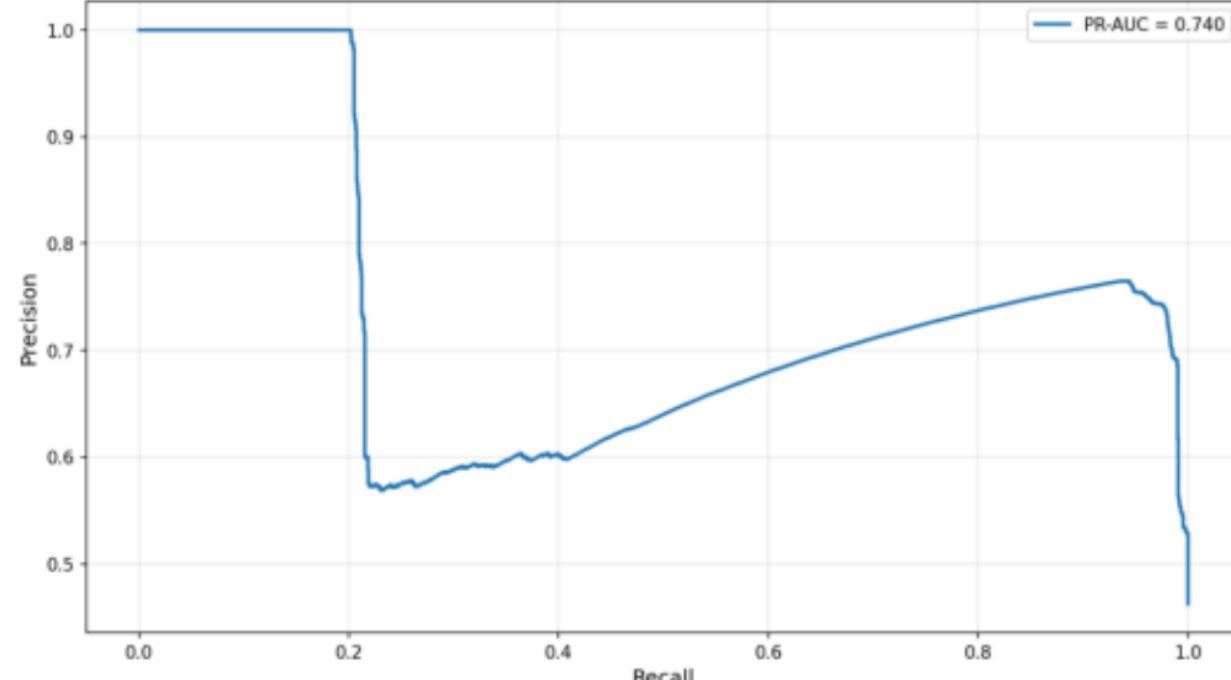
Baseline Model

Baseline Model - Precision-Recall Curve



Logistic Regression Model

Logistic Regression - Precision-Recall Curve



Model Evaluation Summary

Key Findings:

Primary Metric: PR-AUC (Precision-Recall Area Under Curve)

Baseline Model (Z-Score):

- PR-AUC: 0.9997
- F1-Score: 0.9595
- Precision: 0.9995
- Recall: 0.9226

Logistic Regression Model:

- PR-AUC: 0.7398
- F1-Score: 0.3309
- Precision: 0.7157
- Recall: 0.2152

Performance Comparison:

- PR-AUC Improvement: -26.00%
- Better Model: Baseline

Target Performance: PR-AUC ≥ 0.60 (from project requirements)

- Baseline: ✓ MET
- Logistic Regression: ✓ MET

Recommendations:

- The Baseline model shows better performance
- Consider collecting more data if PR-AUC < 0.60
- Monitor for data drift between train and test sets
- Retrain periodically to adapt to market regime changes