Hard Model Comparison Report

Generated: 2025-11-20 12:40:50

1. Hard Dataset Overview

The hard dataset is constructed using a realistic time-based split with a future volatility proxy label (**volatility_spike_future**). Training and test windows are separated chronologically to mimic production deployment.

Split	Rows	Features (numeric)	Spike rate
TRAIN	788	42	2.54%
TEST	198	42	15.15%

2. Hard Model Leaderboard (AUC as Tie-Breaker)

All models below were evaluated on the same HARD test set. The official winner is selected using AUC (Area Under the ROC Curve) as the primary metric, with F1 used as a secondary tiebreaker.

Model	Threshold	Accuracy	AUC	Precision	Recall	F1
random_forest_hard_20251119_114950	0.260	0.8485	0.8346	0.0000	0.0000	0.0000
gradient_boosting_hard_20251119_114951	0.010	0.8384	0.6545	0.0000	0.0000	0.0000
xgboost_hard_20251119_114951	0.260	0.8485	0.5637	0.0000	0.0000	0.0000
logistic_regression_hard_20251119_114950	0.880	0.8485	0.5000	0.0000	0.0000	0.0000

3. Selected Winner (AUC-Based)

Based on the HARD test set, the top-performing model by **AUC** is **random_forest_hard_20251119_114950**. It achieves:
• AUC = **0.8346**
• F1 = **0.0000**
• Accuracy = **0.8485**

This model is selected as the final winner because AUC captures the model's ability to rank volatility spike risk across all thresholds, which is especially important for imbalanced, event-driven problems like crypto spike prediction.

4. Features Used by HARD Models

All four HARD models (Logistic Regression, Random Forest, Gradient Boosting, XGBoost) were trained on the same engineered feature set derived from order book levels, short–/medium–horizon returns, and activity statistics. The winner **random_forest_hard_20251119_114950** was trained on **42** numeric features. The list below shows the exact feature names passed into the model.

Feature 1	Feature 2	Feature 3
ask	best_ask	best_bid
bid	intensity_120s	intensity_30s
intensity_60s	label	last_size
mid_return	midprice	price
price_mean_300s	price_mean_30s	price_mean_60s
price_pct_change	price_std_300s	price_std_30s
price_std_60s	return_10s	return_30s
return_60s	return_max_300s	return_max_30s
return_max_60s	return_mean_300s	return_mean_30s
return_mean_60s	return_min_300s	return_min_30s
return_min_60s	return_std_300s	return_std_30s
return_std_60s	sequence	spread
spread_bps	tick_count_300s	tick_count_30s
tick_count_60s	trade_id	trade_intensity

End of report.