

# n Tree (Grid, Refit by ROC-AUC) — VLO Next-Day D

Generated: 2026-02-25 20:49:35

TRAIN period:  $\leq$  2022-12-31 | TEST period:  $\geq$  2023-01-01

TRAIN size: 2,672 | TEST size: 787

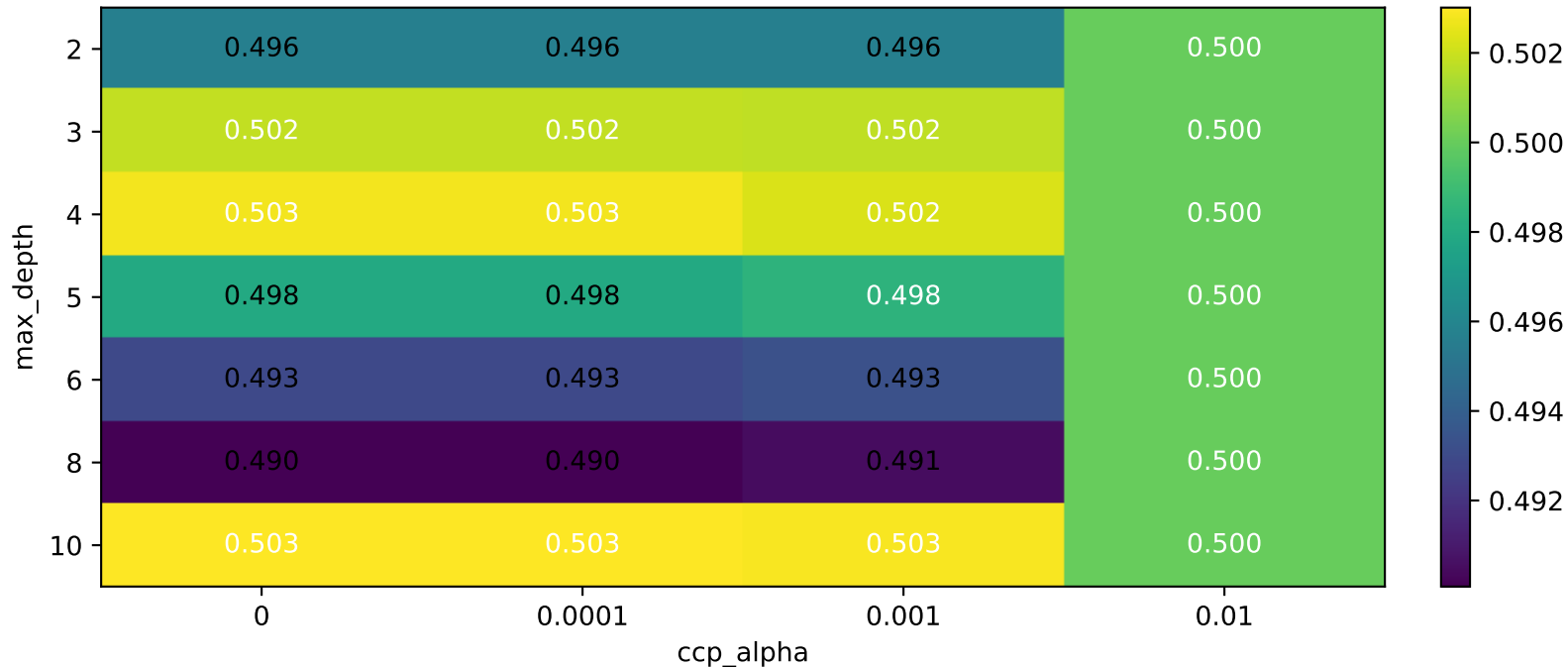
Features — scaled (float + VOL\_\*): 72 | passthrough: 5

= [2, 3, 4, 5, 6, 8, 10, None], min\_split=[2, 5, 10, 25, 50], min\_leaf=[1, 2, 5, 10, 25], ccp\_alpha=[0.0, 0.0001, 0

st params (CV): {'clf\_\_ccp\_alpha': 0.0, 'clf\_\_max\_depth': 10, 'clf\_\_min\_samples\_leaf': 5, 'clf\_\_min\_samples\_split'

Best CV ROC-AUC: 0.5030

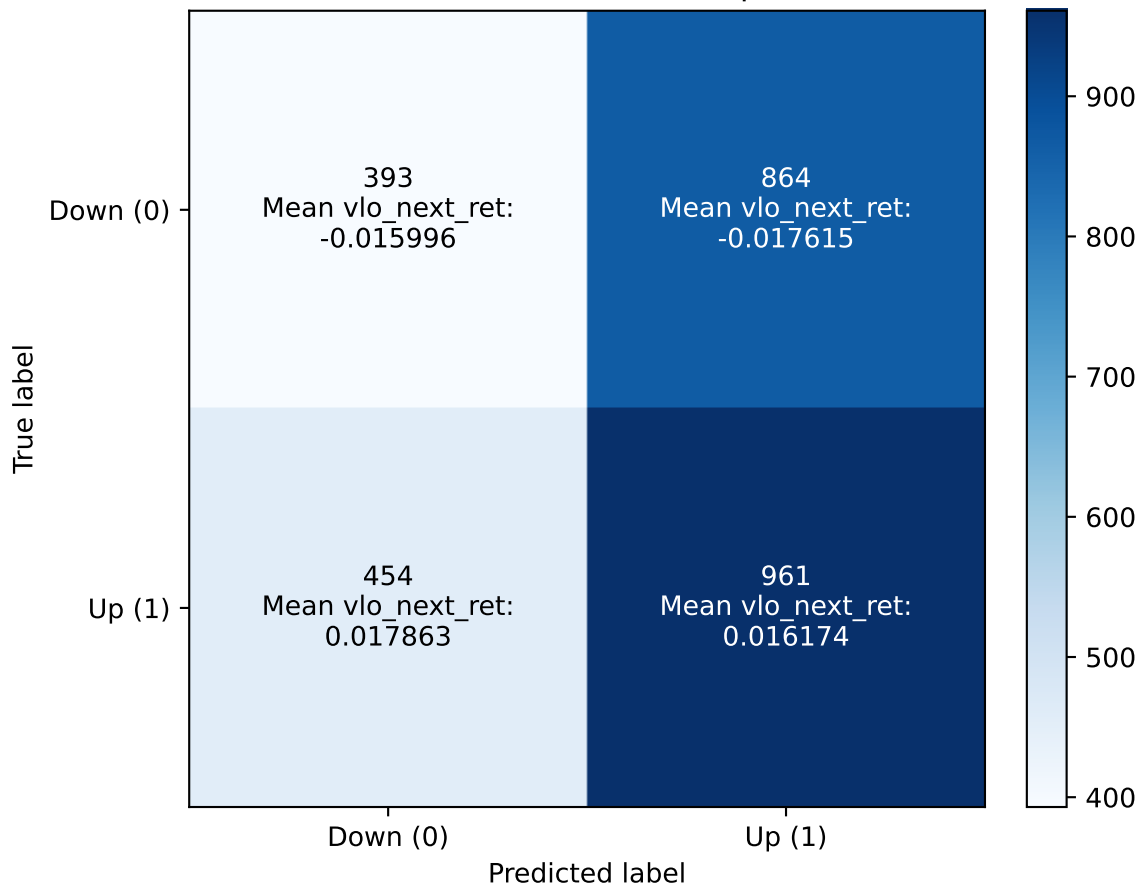
Grid Search — max mean ROC-AUC over (min\_split, min\_leaf)



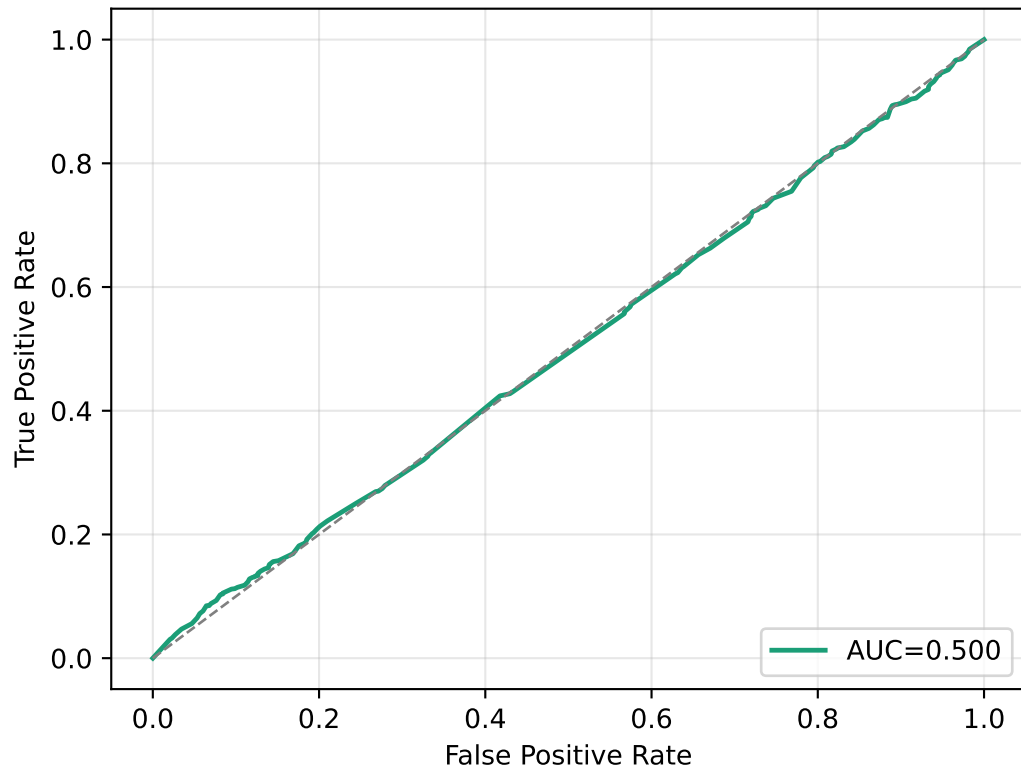
## Cross-Validation ROC-AUC — Training Set ( $\leq 2022-12-31$ )

metric	value
mean ROC-AUC (CV folds)	0.5030
std ROC-AUC (CV folds)	0.0226

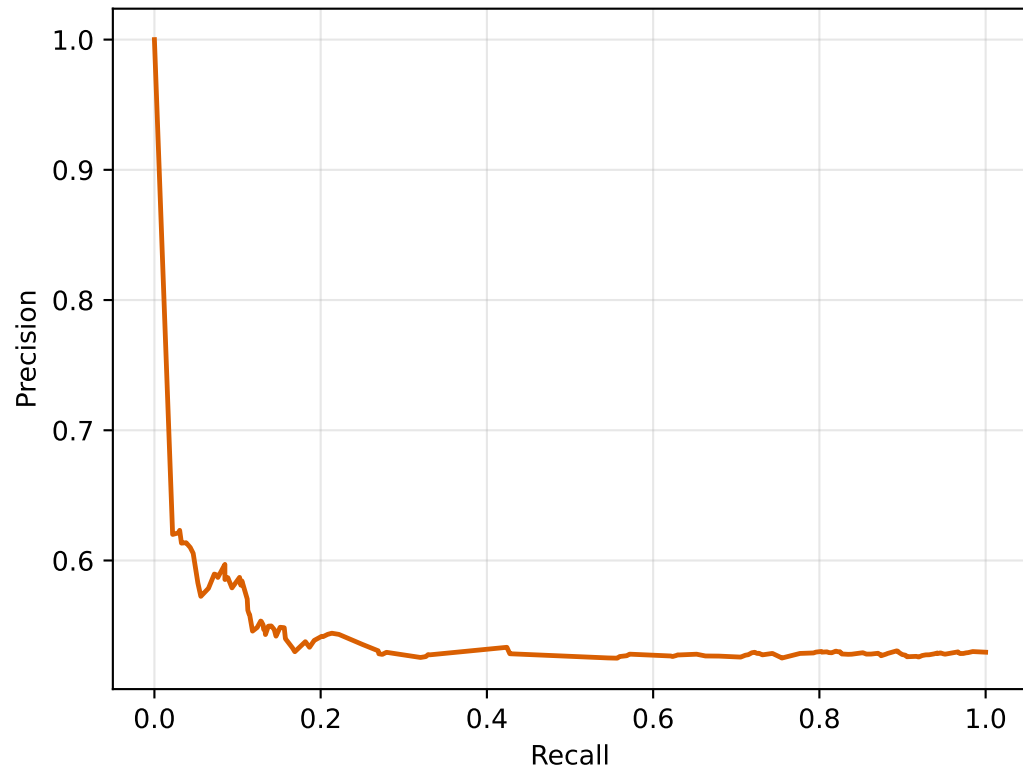
Confusion Matrix — TRAIN (CV predictions)



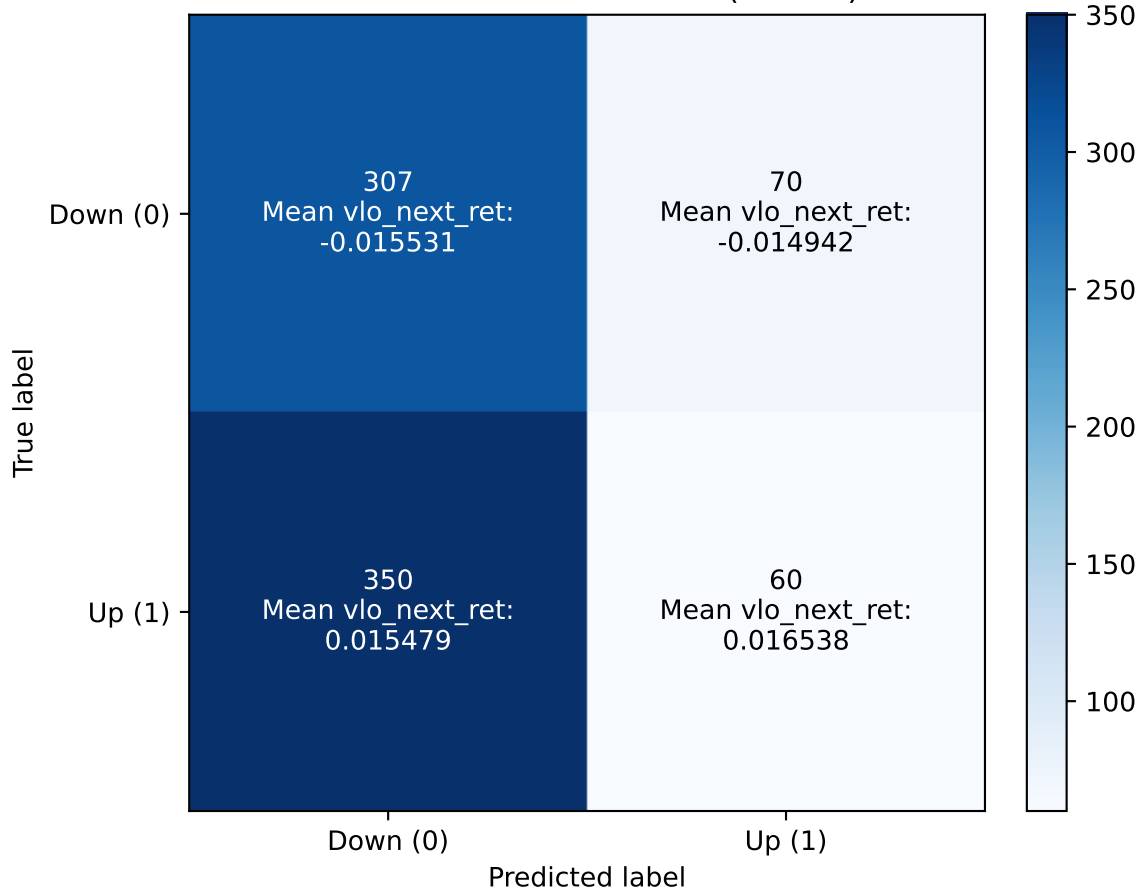
ROC Curve — TRAIN (CV pooled)



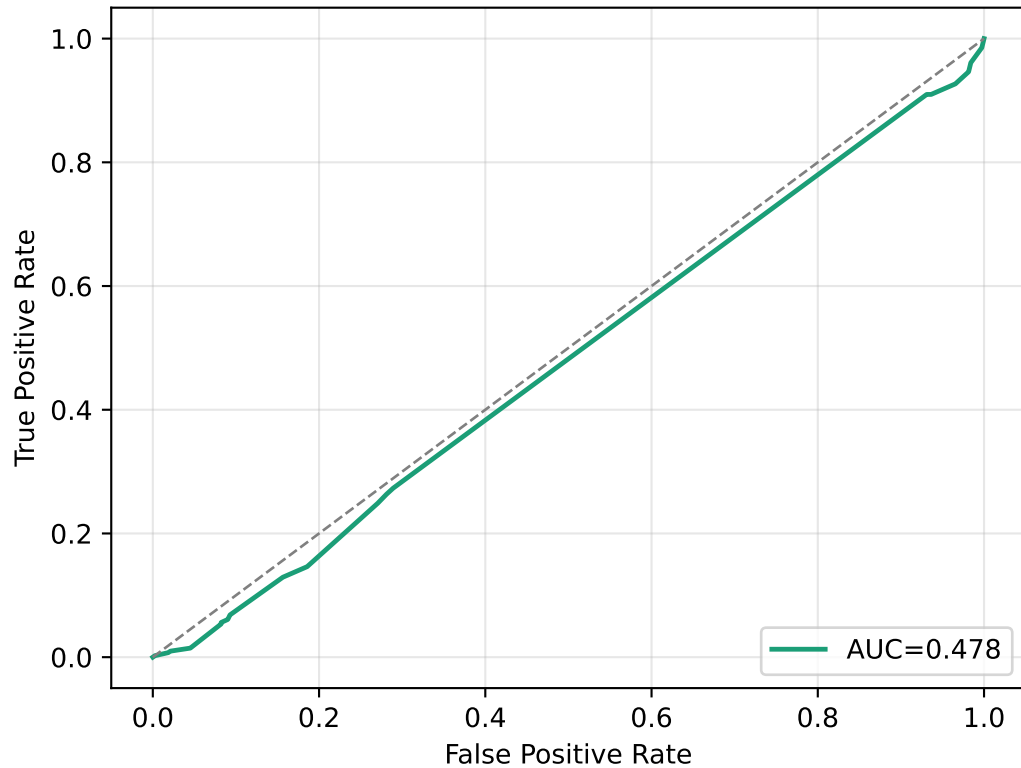
Precision-Recall — TRAIN (CV pooled)



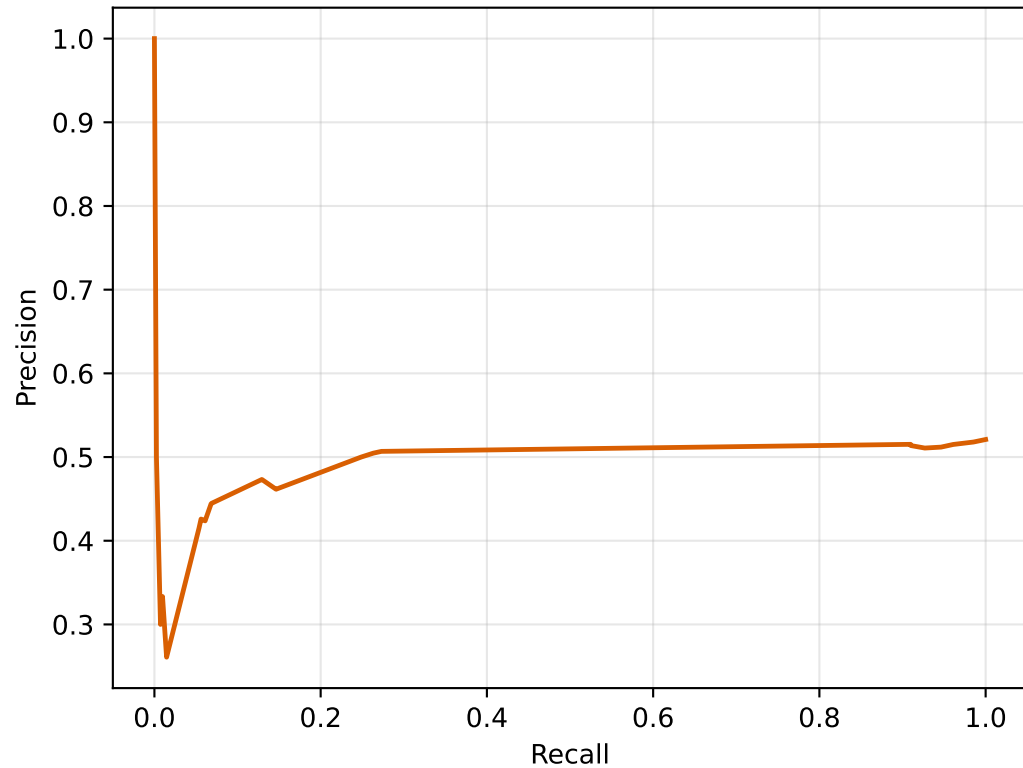
Confusion Matrix — TEST (2023+)



ROC Curve — TEST (2023+)



Precision-Recall — TEST (2023+)

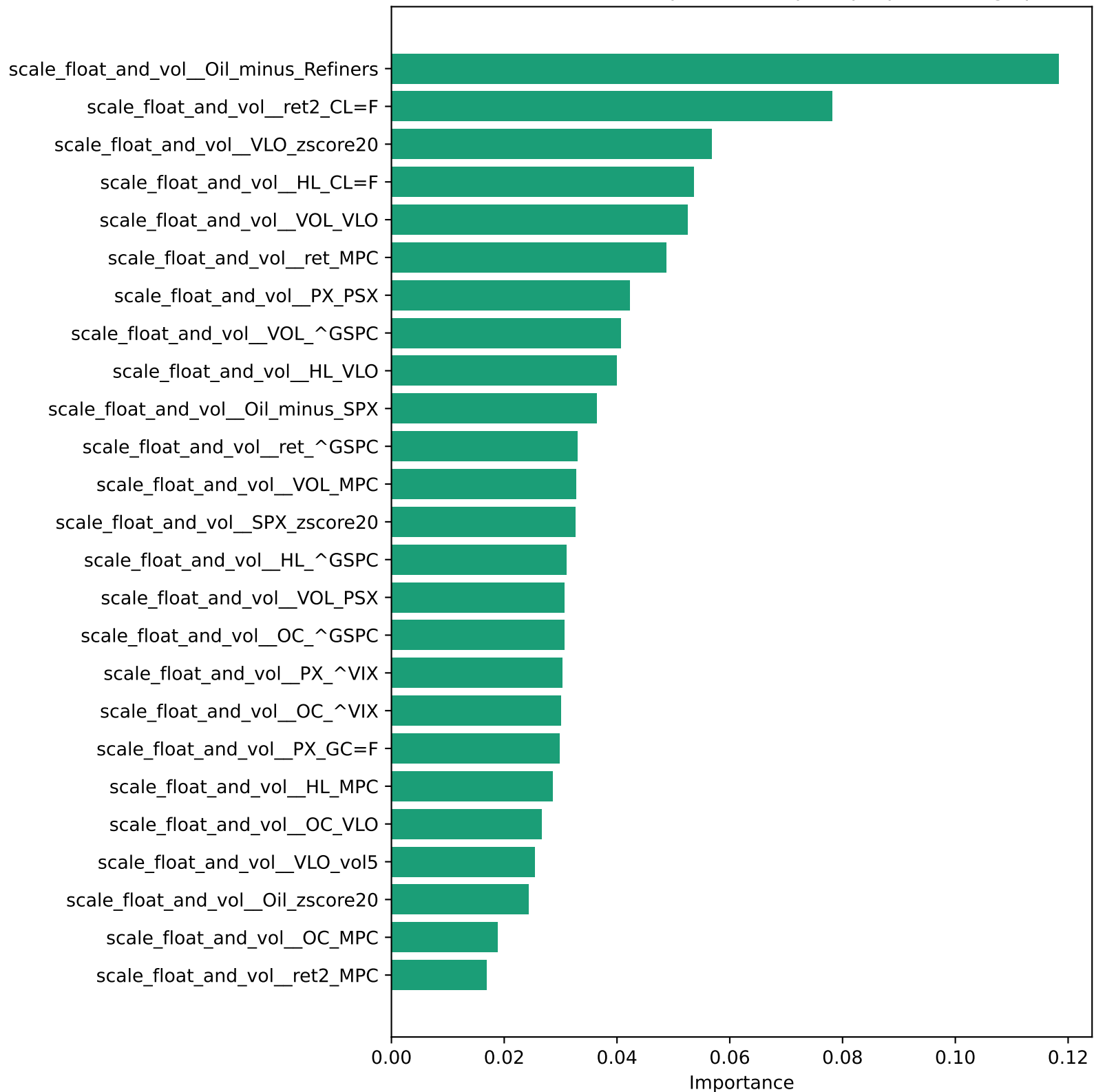


**Classification Report — TEST (2023+)**

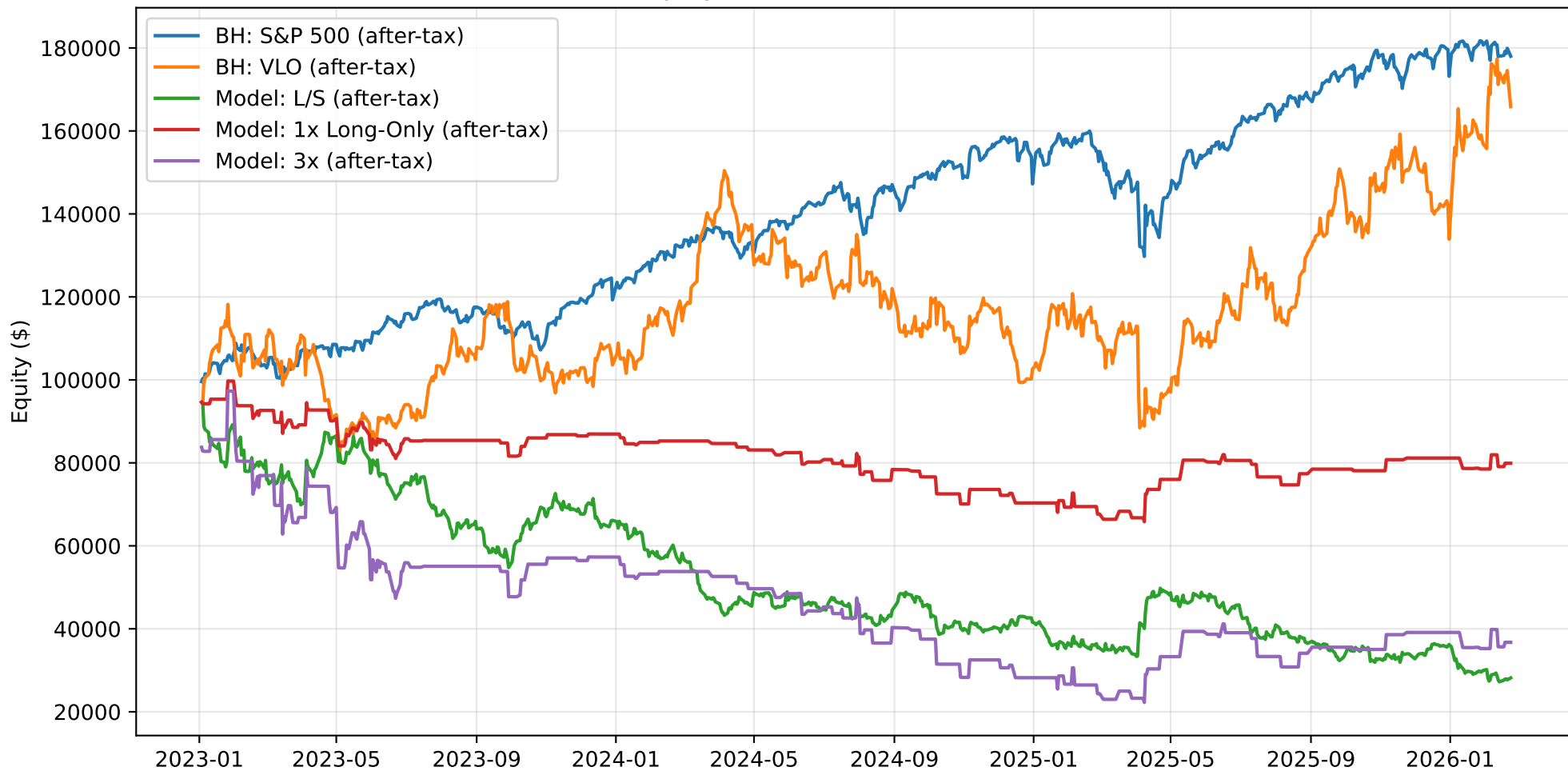
	precision	recall	f1-score	support
0	0.467	0.814	0.594	377
1	0.462	0.146	0.222	410
accuracy			0.466	787
macro avg	0.464	0.480	0.408	787
weighted avg	0.464	0.466	0.400	787



Decision Tree Feature Importances (post-preprocessing space)



After-Tax Equity Curves — TEST: 2023 to most recent



Annual After-Tax Performance (2023+), incl. Average

	BH: S&P 500 (after-tax)	BH: VLO (after-tax)	Model: L/S (after-tax)	Model: 1x Long-Only (after-tax)	Model: 3x (after-tax)
2023	19.78%	9.50%	-30.08%	-8.10%	-31.59%
2024	19.21%	-3.45%	-36.82%	-19.10%	-50.76%
2025	13.32%	29.79%	-12.45%	15.39%	38.69%
2026	-0.32%	16.09%	-21.03%	-1.55%	-6.11%
Average	13.00%	12.98%	-25.10%	-3.34%	-12.44%

## Run Metadata & Time-Split Summary

Report saved: /mnt/batch/tasks/shared/LS\_root/mounts/clusters/dvandyk1/code/Users/dva  
Scaler saved: /mnt/batch/tasks/shared/LS\_root/mounts/clusters/dvandyk1/code/Users/dva  
Model saved: /mnt/batch/tasks/shared/LS\_root/mounts/clusters/dvandyk1/code/Users/dva

TRAIN period:  $\leq$  2022-12-31 | TEST period:  $\geq$  2023-01-01

TRAIN size: 2,672 | TEST size: 787

CV folds: 5

Scaled features (float + VOL \*): 72

Passthrough numeric features: 5

Grid best params: {'clf\_\_ccp\_alpha': 0.0, 'clf\_\_max\_depth': 10, 'clf\_\_min\_samples\_lea

Grid best CV ROC-AUC (refit metric): 0.5030

TEST ROC-AUC: 0.4785 | TEST Accuracy: 0.4663

Backtest assumptions (TEST period only, AFTER-TAX curves shown):

Initial capital: \$100,000

Tx cost per side: 2.00 bps

Taxes (annual on positive P&L, loss carry-forward):

- Long-term (buy & hold): 20.0%

- Short-term (active): 37.0%

Short borrow APR (L/S shorts): 0.30%

3x long-only daily cost (financing+expense): 0.0595%

Stop-loss: NOT APPLIED

Feature exclusions (pre-model): vlo\_next\_ret, vlo\_next\_dir