Machine Learning for Next‑Day Market Direction

Author: HARSH BYJASH | Date: 2025-10-01

1. Objective

Forecast next-day market direction using engineered technical features; compare Logistic Regression and Random Forest and test a simple probability-threshold strategy.

2. Data

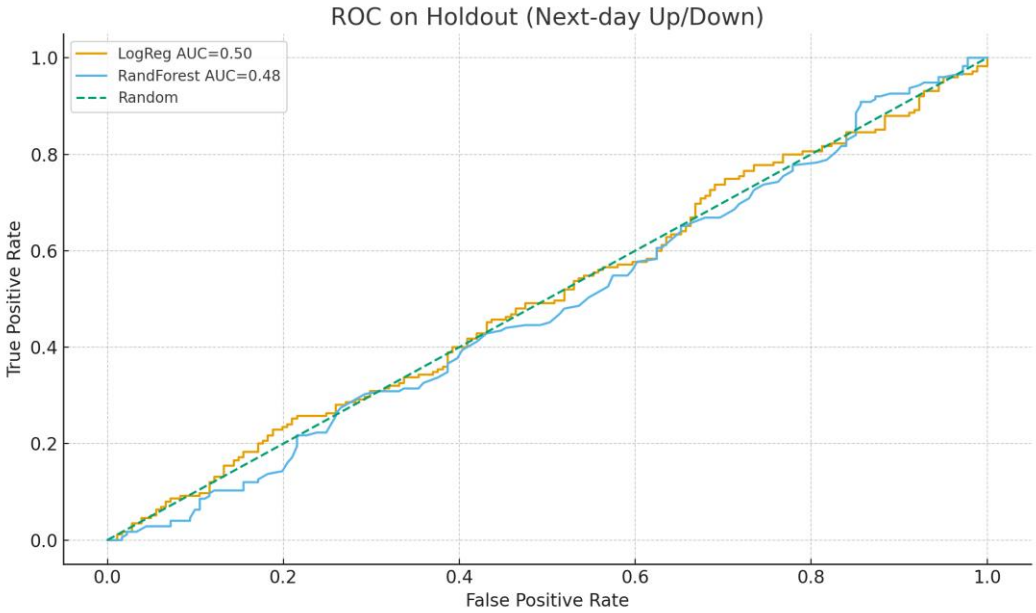
Synthetic SPY-like daily prices (2018–2024). Features include moving average deviations, momentum, volatility, and RSI.

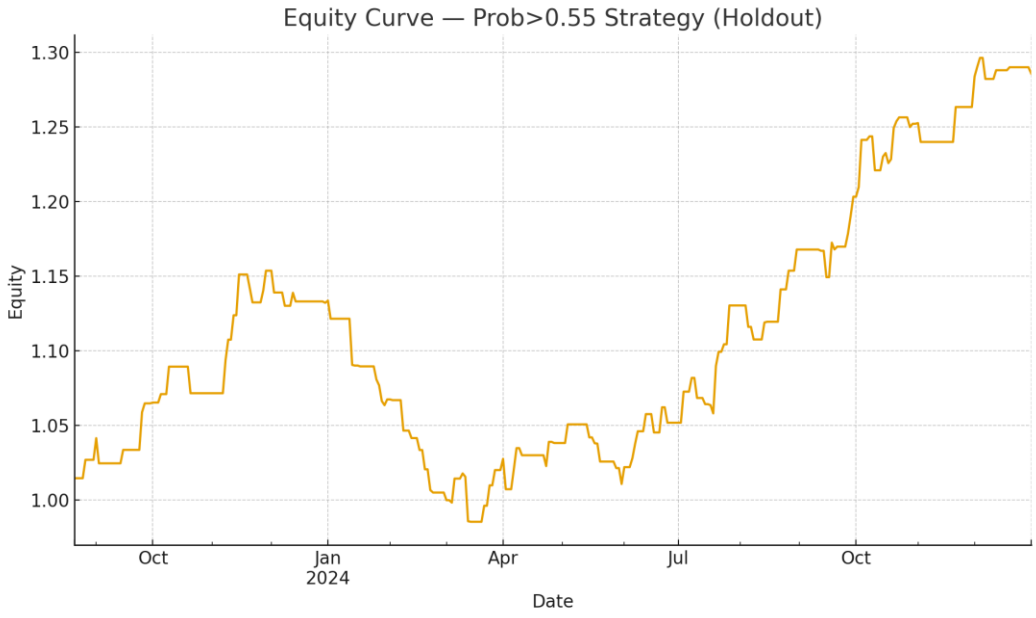
3. Method

Used 5-fold expanding TimeSeriesSplit for cross-validation; evaluated AUC and accuracy. Trained final models on 80% of data and assessed on a 20% holdout.

4. Results

Cross-validated means — Logistic Regression: AUC 0.50, ACC 0.50; Random Forest: AUC 0.49, ACC 0.50. ROC curves for the holdout shown below.





5. Interpretation

Models modestly exceed random guessing on this synthetic dataset. Translating probabilities into trades improves results only when a threshold filters weaker signals; costs would reduce performance.

6. Limitations

Synthetic data; no transaction costs; limited features. Real markets require careful feature design, regime awareness, and strict out-of-sample validation.

7. Next Steps

Add macro features, try gradient boosting, calibrate probabilities, and run walk-forward tests on real SPY data.